UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Neven Tomašić

DESIGNING THE CHEMICAL COMPOSITION OF STEEL WITH THE REQUIRED HARDENABILITY USING COMPUTATIONAL METHODS

DOCTORAL THESIS

Rijeka, 2025.

UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Neven Tomašić

DESIGNING THE CHEMICAL COMPOSITION OF STEEL WITH THE REQUIRED HARDENABILITY USING COMPUTATIONAL METHODS

DOCTORAL THESIS

Supervisor: Assoc. Prof. D. Sc. Dario Iljkić

Co-supervisor: Assoc. Prof. D. Sc. Wojciech Sitek

Rijeka, 2025.

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET

Neven Tomašić

DIZAJNIRANJE KEMIJSKOG SASTAVA ČELIKA PREMA TRAŽENOJ PROKALJIVOSTI RAČUNALNIM METODAMA

DOKTORSKI RAD

Mentor: Izv. prof. dr. sc. Dario Iljkić

Komentor: Izv. prof. dr. sc. Wojciech Sitek

Rijeka, 2025.

Doctoral thesis supervisor: Assoc. Prof. D. Sc. Dario Iljkić, University of Rijeka, Faculty of Engineering, Croatia

Doctoral thesis co-supervisor: Assoc. Prof. D. Sc. Wojciech Sitek, Silesian University of Technology, Scientific and Didactic Laboratory of Nanotechnology and Materials Technologies, Poland

The doctoral thesis was defended on _____ at the University of Rijeka, Faculty of Engineering, Croatia, in front of the following Evaluation Committee:

1. Assoc. Prof. D. Sc. Sunčana Smokvina Hanza, University of Rijeka, Faculty of Engineering, Croatia - Committee Chair
2. Assist. Prof. D. Sc. Matej Fonović, University of Rijeka, Faculty of Engineering, Croatia
3. Prof. D. Sc. Darko Landek, University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture

We're smart enough to invent an AI,

dumb enough to need it,

and so stupid we can't figure out if we did the right thing.

Jerry Seinfeld

ACKNOWLEDGMENTS

The journey that began long ago has now reached a significant milestone. Along the way, I have had the privilege of meeting and becoming familiar with numerous kind, positive, and inspiring individuals. Many of them—if not all—have taught, guided, and directed me to where I stand today. A single book would not be enough to express my gratitude to them all.

A special salute goes to my colleagues from the Department of Materials Engineering: Sunčana Smokvina Hanza, Dario Iljkić, Matej Fonović, and our dear Loreta Pomenić and Božo Smoljan. They have been with me practically from day one, when I first stepped into the department—a place that was still evolving into what it is today: the Department of Materials Engineering at the Faculty of Engineering.

Life would be dull without its little miracles. One day, a dear friend of mine asked another person, who is also very dear to me, a simple yet pivotal question, and not for himself: *Professor, do you have an interesting research topic that could one day develop into a doctoral thesis?* This conversation took place over 1,000 kilometers away from me and without my knowledge. The answer was positive, and two years later, I have the privilege of expressing my gratitude on these pages to both of them. Thank you, my friend Lovro Liverić, for finding the right words to ask that question. And a very special thanks to the person who gave the right answer and, after nearly 20 years of knowing each other, reserved a topic for me and dedicated his invaluable time to becoming my mentor—dear Wojciech Sitek. Before that meeting, a few small miracles had already taken place between us, foreshadowing the start of this exciting and enlightening journey—but I'll keep those memories to myself. Although I have already acknowledged him, I must once again emphasize my appreciation for his scientific criticism, and above all, his sincere dedication and willingness to help. His vast knowledge played a crucial role in bringing this thesis to its final stage. Once again, big thanks to Dario.

The journey continues … *Author*

ABSTRACT

The objective of this research was to develop a predictive model using artificial neural networks to determine the optimal chemical composition to achieve required hardenability in steel. The chemical composition of steel has great influence on its hardenability; however, the contribution of alloying elements is not purely additive or linear. The correlation between hardness values at various distances from the quenched end of the Jominy specimen and the chemical composition of steel, based on current knowledge of the subject is impossible mathematically defined. Since hardness is affected by both microstructure and chemical composition, microstructural data was also incorporated into the research. Supervised artificial neural networks were employed to create a sophisticated regression models which are based on a representative dataset of steels for heat treatment.

A representative dataset of 470 steel samples was collected, with each sample including Jominy test results and chemical compositions. Additionally, an indication of significant martensite presence in the microstructure was included in the input data. Four distinct approaches to dataset representation were investigated. Each new approach offered ANN models with better results and more options for future research.

The developed models were experimentally validated, demonstrating its effectiveness. This innovative approach enables automated and precise prediction of chemical composition based on the required Jominy curve (hardenability), offering significant benefits to the mechanical engineering and manufacturing industries.

Keywords: Heat Treatment, Hardenability, Chemical compositions, Microstructure, Jominy curve, Deep learning

PROŠIRENI SAŽETAK

Cilj istraživanja u ovom doktorskom radu bio je razviti model koji predviđa optimalni kemijski sastava čelika potreban za postizanje zahtijevane prokaljivosti pomoću umjetnih neuronskih mreža. Kemijski sastav čelika utječe na prokaljivosti. Međutim, utjecaj legirajućih elemenata nije strogo aditivan niti linearan. Korelacija između vrijednosti tvrdoće na različitim udaljenostima od gašenog čela Jominyjevog uzorka i kemijskog sastava čelika je složena i ne može se precizno definirati tradicionalnim matematičkim metodama.

Prikupljen je reprezentativni skup podataka od 470 uzoraka čelika, pri čemu svaki uzorak uključuje rezultate Jominyjeva testa (tvrdoću izražena u HRC, na 13 udaljenosti od gašenog čela) i kemijski sastav (maseni udjeli ugljika, mangana, silicija, kroma, nikla, molibdena i bakra). Budući da na tvrdoću utječu i mikrostruktura i kemijski sastav, mikrostrukturni podaci također su uključeni u istraživanje. U ulazne podatke uključen je i pokazatelj prisutnosti martenzita u mikrostrukturi. Pokazatelj značajne prisutnosti martnezita definiran je s prisutnosti od 50% martenzita u mikrostrukturi.

Nadzirane umjetne neuronske mreže korištene su za razvoj sofisticiranih regresijskih modela temeljenih na reprezentativnom skupu podataka čelika za toplinsku obradu. Istražena su četiri različita pristupa predstavljanju skupa podataka, pri čemu je svaki novi pristup omogućio poboljšane modele umjetnih neuronskih mreža i otvorio nove mogućnosti za buduća istraživanja.

Za modeliranje je korišten program MATLAB 2023b, uključujući aplikacije Neural Network Fitting i Experiment Manager. Najoptimalniji modeli odabrani su na temelju najnižeg parametra korijena srednje kvadratne pogreške (engleski *root mean squared error* - RMSE).

Za eksperimentalnu validaciju modela korišteno je pet čelika različitih prokaljivosti. Predviđeni maseni udjeli kemijskih elemenata za svih pet čelika nalazili su se unutar granica definiranih klasama čelika. Razvijeni modeli eksperimentalno su potvrđeni,

čime je dokazana njihova učinkovitost. Ovaj inovativni pristup omogućuje automatizirano i precizno predviđanje kemijskog sastava na temelju zahtijevane Jominyjeve krivulje (prokaljivosti), nudeći značajne prednosti za strojarstvo i proizvodnu industriju.

Ključne riječi: toplinska obrada, prokaljivost, kemijski sastav, mikrostruktura, Jominyjeva krivulja, duboko učenje

Table of Contents

1. INTRODUCTION

1.1. Origin of work

The selection of steel for specific machine parts or constructions begins with defining the required properties. Among these properties, hardenability of steel is one of the most critical criteria for steels intended for quenching. Hardenability of steel is the ability of steel to achieve martensitic structure in deeper layer. It depends on the chemical composition of steel.

The most commonly used method for characterizing hardenability is the Jominy-Boegehold test, which produces a Jominy curve, which is a graphical representation of hardness values along the length of a test specimen. Hardenability is influenced by the prior austenite grain size, cooling rate, and the kinetics of austenite transformation. The contribution of alloying elements is crucial and it is not purely additive or linear. The correlation between hardness values at various distances from the quenched end of the Jominy specimen and the chemical composition of steel, based on current knowledge, cannot be fully explained using mathematical formulas. Microstructure significantly influences hardness and must be evaluated in conjunction with chemical composition.

Consequently, designing the chemical composition of steel to achieve a desired hardenability, defined by the shape of the Jominy curve, is a highly complex task. This process is difficult to accomplish without the aid of computational methods. Computational modelling is a relatively inexpensive and efficient method widely used in materials science. Nowadays, with the advancement of computers and available software, artificial neural networks are used increasingly. Deep learning employs methods that enable computers to learn from real datasets by modeling nonlinear correlations between material properties and influencing factors.

In the modern era of optimization, achieving specific material properties at the lowest possible cost without compromising quality is crucial. Therefore, there is a growing need to develop steel with an optimized chemical composition that not only meets the required hardenability but also minimizes production costs.

1.2. Hypothesis and research objective

The main hypothesis of this doctoral thesis posits that it is possible to design chemical composition of steel with the required hardenability. The underlying premise is that the relationship between chemical composition, on one side and microstructure and hardenability on other side, can be modeled effectively to achieve desired outcomes.

The research objective stems from this hypothesis: to develop a predictive model based on artificial neural networks that can determine the optimal/suitable chemical composition required for achieving a specific hardenability of steel. This model incorporates the effects of both chemical composition and microstructural characteristics on steel hardness.

A representative dataset, comprising Jominy test results, mass fractions of seven chemical elements, and microstructural data, is utilized to establish and quantify the complex relationships that govern hardenability.

The expected outcome of this research is a model that facilitates the calculation of the optimal/suitable chemical composition of steel to achieve specific hardenability of steel, i.e. Jominy curve profile.

This hypothesis driven approach aims to provide an innovative framework for designing steel compositions that meet precise hardenability criteria, with potential applications in advancing material science and industrial steel production.

1.3. Research methodology and scientific contribution

To achieve the objectives outlined in the hypothesis and research, the first critical step was to establish of a representative dataset, containing the chemical compositions of steels and their corresponding experimentally obtained Jominy test results. A comprehensive dataset comprising 470 steels for quenching and tempering, and case hardening by carburizing steels, with hardness data from the Jominy-Boegehold test (Jominy curve) serving as input predictors. Data were collected at 13 distinct distances from the quenched end, while the chemical composition was used as output response data for the same steels. The modeling process utilized mass fractions of carbon, silicon, manganese, chromium, nickel, molybdenum, and copper. Recognizing the role of microstructure in influencing hardness values, microstructure was also incorporated into the model, providing a more nuanced understanding of the relationship between chemical composition on one side and hardness and microstructure on other side.

For the advanced regression analysis necessary to determine the relationships between the Jominy test results (hardness and microstructure) and the chemical composition of steel, Artificial Neural Networks (ANNs) in MATLAB 2023b were employed. Different models were generated for various data organizations, with the Regression Learner App used to train the neural network models. To prevent overfitting, 10-fold cross-validation was implemented during training. The model development process, encompassing both the training and testing phases. Models with the lowest test root mean square error (RMSE) were selected as the most accurate.

In this thesis, a novel and innovative approach was developed by enabling the automated and highly accurate prediction of the chemical composition of steel based on the required hardenability. This method uniquely incorporates the consideration of microstructural variations at different distances from the quenched end of the

Jominy specimen, representing a significant advancement in the field of material science and steel processing.

1.4. Structure of the doctoral thesis

The doctoral thesis is organized into four main sections. The first section introduces the hypothesis, establishes the scientific significance of the study, and outlines the methodology employed. It serves as the foundation for understanding the research objectives and approach, while providing an overview of the thesis structure.

The second section delves deeper into the theoretical background and includes a comprehensive literature review, divided into two major parts. The first part is concerned with the transformations of undercooled austenite and its relevance to hardenability. The historical development of hardenability characterization is traced, beginning with early studies and culminating in modern computational modeling techniques. A detailed discussion of the Jominy-Boegehold method is included to illustrate one of the most widely used characterization methods, along with its result—the Jominy curve. The second part explores the evolution of neural networks, beginning with early efforts to simulate biological neurons. It highlights seminal contributions, such as the foundational work of McCulloch and Pitts, Rosenblatt's perceptron, and Widrow and Hoff's formalization of learning through quantitative error measures. Key concepts, including the architecture of feedforward neural networks, backpropagation learning, and non-linear activation functions like sigmoid, ReLU, and Leaky ReLU, are presented to provide a comprehensive understanding of neural network theory.

The third section focuses on analytical and experimental research, offering a detailed explanation of the methodology applied in the study. It presents the findings systematically, supported by visual data representations such as graphs and tables. The models are analyzed and interpreted in the context of the theoretical background

established earlier, demonstrating the correlation between chemical composition and the Jominy curve.

The final section provides a conclusion and critical evaluation of the research. It summarizes the key findings, identifies the strengths and limitations of the study, and discusses its broader implications. The experiment that validated the numerical models is also explained. The thesis concludes by proposing potential avenues for future research, emphasizing how the findings could be applied to address related challenges or explore new directions in the field.

2. THEORETICAL BACKGROUND AND REVIEW OF PREVIOUS STUDIES

2.1. Undercooled austenite transformation

Steel is the most widely used construction material among all metals, due to its cost, the availability of iron, and its unique combination of properties. Another feature that contributes to steel's importance as a construction material is its ability to be modified and enhanced to meet specific requirements.

Heat treatment facilitates the modification of steel's properties by intentional alteration of its microstructure. The alteration of microstructure during these processes influences the mechanical properties, particularly hardness and strength. During heat treatment, it is essential to define the appropriate heating temperature for the steel specimen, as well as the cooling rate. A controlled cooling rate is critical for achieving the desired hardness and strength, while also minimizing undesirable effects such as brittleness, distortion, or excessive residual stresses.

The transformation temperatures of steel are critical temperatures at which alterations in the crystal structure occur, mostly influenced by the carbon content within the microstructure. In addition to carbon, other chemical elements also influence the transformation temperatures of steel [1]. An especially interesting temperature is the critical temperature of austenite (γ - Fe) formation during heating and completely its transformation upon cooling which is denoted A_1. For plain carbon steels, A_1 is approximately 725 °C. Results of austenite decomposition during cooling below 725 °C can include pearlite, bainite, or martensite, depending on the cooling conditions [2].

The driving force for a phase transformation is the difference in Gibbs free energy between the initial and final phases [3]. A system tends to undergo a phase

transformation in the direction that decreases its Gibbs free energy, moving toward a more stable phase.

$$\Delta G = G_{\text{final}} - G_{\text{initial}} \quad (2.1).$$

When $\Delta G < 0$, the system will naturally transform from the initial phase to the final phase to reach a lower energy, more stable state.

For phase transformations, equation 2.1 can be expressed in terms of enthalpy, temperature, and entropy:

$$\Delta G = \Delta H - T\Delta S \quad (2.2),$$

where H is enthalpy, T is absolute temperature and S is entropy.

Austenite has a face-centered cubic (FCC) crystal structure, while ferrite (α - Fe) has a body-centered cubic (BCC) crystal structure. Austenite FCC structure contains a greater number of iron atoms per unit cell than ferrite BCC structure. Austenite FCC structure contains four iron atoms, while ferrite BCC structure contains only two iron atoms. Furthermore, austenite FCC structure provides an increased number of available slip planes. This atomic configuration facilitates increased atomic mobility and randomness. Pearlite is a two-phase mixture of ferrite and cementite (Fe_3C). The alternating layers of ferrite and cementite introduce structural regularity, further reducing atomic randomness.

Even if austenite has a higher enthalpy (H) compared to pearlite, its higher entropy (S) makes Gibbs free energy for austenite lower at elevated temperatures, stabilizing the phase (Figure 2.1).

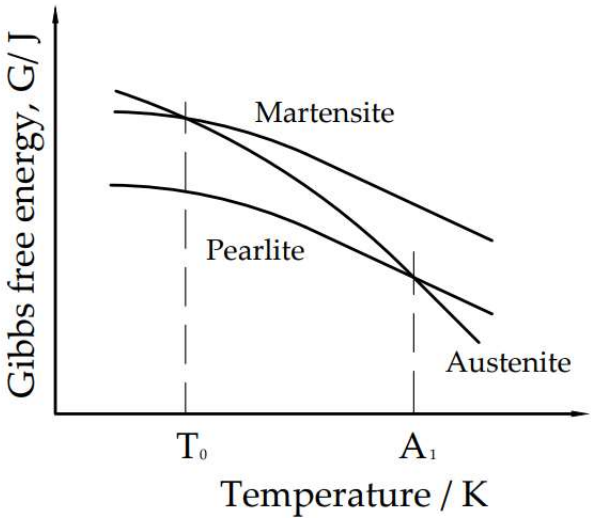


Figure 2.1. Gibbs free energy vs. temperature diagram [4]

Austenite is unstable below A_1 and it tends to transform to other phases depending on cooling conditions. During cooling, the transformation to martensite can only occur if the steel is first heated to the austenitization temperature. If pearlite is present instead of austenite before cooling, martensite will not form, as pearlite is already thermodynamically stable.

The critical temperatures are dependent upon whether the steel is subjected to heating or cooling. Critical temperatures for starting and finishing the transformation to austenite during heating are denoted, respectively, by A_{c1} and A_{c3} for hypo-eutectoid steels (with wt. % of carbon less than 0.8) and by A_{c1} and A_{ccm} for hypereutectoid steels (with wt. % of carbon between 0.8 and 2). These temperatures are higher than the corresponding critical temperatures for the start and finishing of the transformation from austenite during cooling, which are represented by A_{r3} and A_{r1} for hypoeutectoid steels and by A_{rcm} and A_{r1} for hypereutectoid steels with carbon content higher than 0.8%).[1] The faster the heating, the higher is the A_c temperature; the faster the cooling,

[1] Suffix '*c*' comes from French word for heating (*Chauffage*) and '*r*' comes from French word for cooling (*Refroidissement*).

the lower is the A_r temperature. As heating and cooling rates become endlessly slow, these critical temperatures converge on the equilibrium values A_1, A_3, and A_{cm} (Figure 2.2).

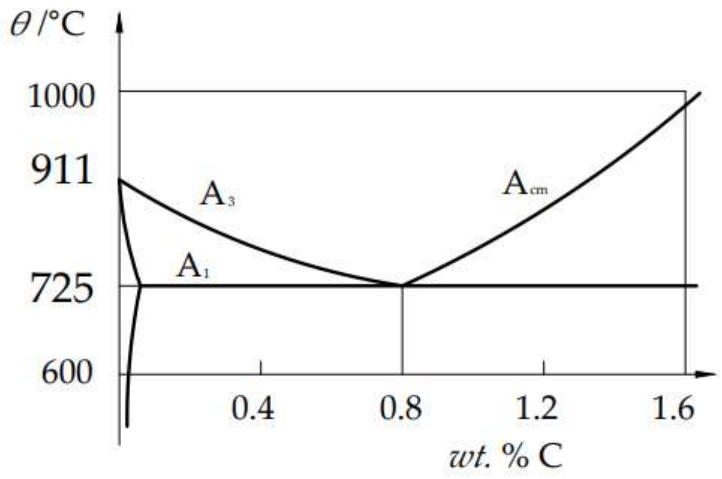


Figure 2.2. The critical temperatures for steels

2.1.1. Pearlite transformation

The undercooling of austenite determines the microstructure that forms during the transformation of austenite. As the hypoeutectoid steel is cooled slightly below the A_3 temperature, proeutectoid α phase begins to form. The formation of ferrite is a diffusion-driven process. As ferrite forms, remained austenite enriches with carbon atoms, because there is no place for carbon atoms in ferrite. Both austenite (with an FCC crystal structure) and ferrite (with a BCC crystal structure) are primarily composed of iron atoms. During the transformation, these iron atoms rearrange to adapt to the new crystal structure. This rearrangement does not require the diffusion of iron atoms across large distances.

The cooling rate plays a significant role in determining both the amount of ferrite that forms and the form it takes. Ferrite can be classified into two types: allotriomorphic and idiomorphic ferrite. Allotriomorphic ferrite is named for its irregular, non-equiaxed shape, as it grows along the prior austenite grain boundaries (Figure 2.3).

Idiomorphic ferrite refers to equiaxed (regular, roughly spherical) ferrite crystals that form within the austenite grain.

During slow cooling, allotriomorphic ferrite forms because the atoms have sufficient time to migrate, allowing ferrite to develop at the grain boundaries of austenite. In contrast, faster cooling rates promote the formation of equiaxed ferrite. This occurs because, at higher cooling rates, atoms may not have enough time to align into the elongated, boundary-following structure of allotriomorphic ferrite, resulting in more uniform ferrite formation within the austenite grain [5].

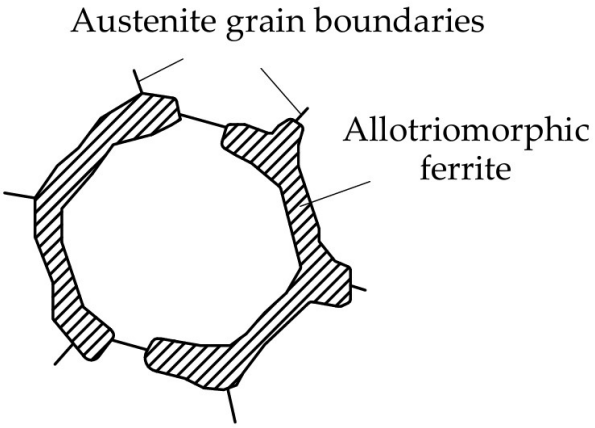


Figure 2.3. Schematic illustration of grain boundary allotriomorphic ferrite formation

The pearlite transformation is a diffusion-based transformation that results in the formation of a two-phase microstructure composed of ferrite and cementite (Fe_3C) arranged in a lamellar (layered) structure. The transformation occurs at temperatures below the A_1 temperature (eutectoid temperature). As the steel was cooled below the A_3 temperature, the formation of ferrite caused the remaining austenite to become enriched with carbon. This process continued until the temperature reached A_1, at which point the remaining austenite had a carbon content of 0.8 wt.% (eutectoid composition). The eutectoid transformation is a solid-state phase change in which single-phase austenite, with an approximate carbon content of 0.8 wt.%, transforms

into a two-phase microstructure. The ferrite transformed before the eutectoid temperature was achieved by cooling is referred to proeutectoid in order to indicate that it has been formed by a mechanism other than the eutectoid transformation.

During the pearlite transformation, carbon atoms begin to diffuse, and the austenite decomposes into alternating layers of ferrite and cementite. Carbon atoms diffuse out of the ferrite regions and into the cementite regions, forming the characteristic lamellar structure. Cementite forms as a stable carbide to accommodate the excess carbon.

The nucleation of pearlite is a heterogeneous process, meaning it typically occurs at sites where the energy barriers for nucleation are lower. The austenite/proeutectoid ferrite boundaries are ideal nucleation sites because the boundary provides an interface that reduces the energy required to form new phases (Figure 2.4).

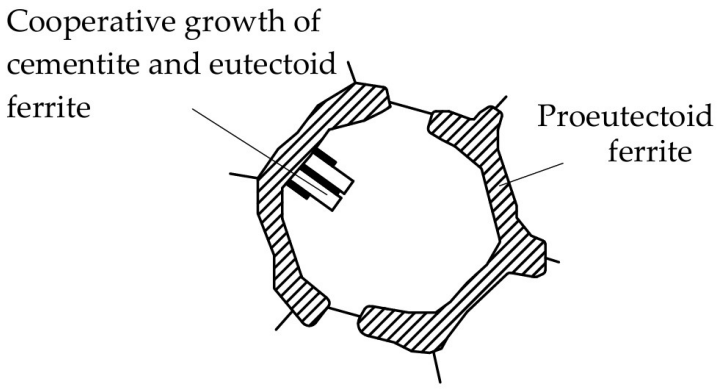


Figure 2.4. Schematic illustration of pearlite formation

Carbon atoms readily diffuse along austenite/proeutectoid ferrite boundaries, facilitating the formation of the cementite component of pearlite. In eutectoid steels, nucleation occurs at austenite grain boundaries due to the absence of proeutectoid ferrite. Conversely, in hypereutectoid steels, nucleation takes place at austenite/proeutectoid cementite boundaries.

The simple calculation based on the lever rule shows that in a pearlite, approximately 88 wt.% is ferrite, while the remaining 12 wt.% is cementite. Ferrite is a soft, ductile, and relatively low-strength phase. Cementite is a hard, brittle, and high-strength phase. Pearlite has a good combination of strength and toughness.

In faster cooling processes, with higher undercooling carbon diffusion is restricted due to less time and lower temperatures. Carbon atoms do not have sufficient time to segregate into distinct regions at a larger scale. Due to the restricted diffusion of carbon, the formation of pearlite occurs with narrower (finer) interlamellar spacing between the ferrite and cementite lamellae [6]. This is because the carbon content changes more quickly, leading to a denser arrangement of the alternating phases.

During slower cooling and lower undercooling (temperature closer to A_1) diffusion of carbon atoms occurs more freely due to more time and higher temperatures, leading to widely spaced (coarser) lamellae (Figure 2.5).

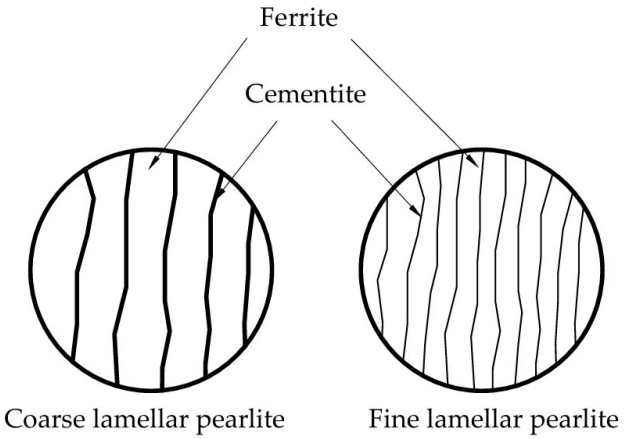


Figure 2.5. Difference in pearlitic structure due to annealing and normalizing

Fine pearlite can be obtained by normalizing, while coarse pearlite can be obtained by annealing (Figure 2.6).

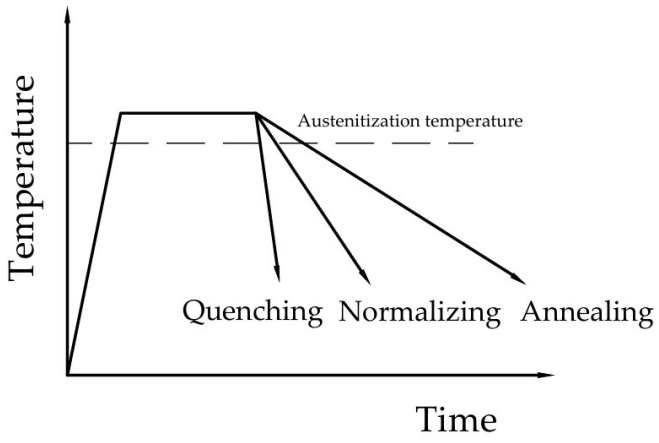


Figure 2.6. Process nomenclature based on cooling rate

2.1.2. Martensitic transformation

Increased undercooling of austenite, achieved through a higher cooling rate (quenching), reduces the time available for atomic diffusion, significantly affecting the resulting microstructure of the steel. At lower temperatures, the displacive decomposition of austenite occurs, where atoms shift positions without long-range diffusion. Transformation occurs by crystallographic shear. This is in contrast to the reconstructive decomposition of austenite (e.g., pearlite transformation), which involves atomic diffusion and occurs at slower cooling rates at higher temperatures.

Quenching prevents carbon atoms from diffusing out of the austenitic lattice in adequate time. To optimize quenching, it is essential to determine the upper critical cooling rate. The upper critical cooling rate is defined as the minimum cooling rate at which diffusion-controlled precipitation during quenching is inhibited, allowing all dissolved alloying element atoms to stay in a supersaturated solid solution. When the cooling rate exceeds the upper critical cooling rate, increased undercooling of austenite occurs, leading to the martensitic transformation. The upper critical cooling rate is the minimum cooling rate required to form a 100% martensitic structure in steel during quenching. Being an athermal reaction, martensitic transformation is only dependent on temperature and not on time. The Martensite start (M_s) and martensite

finish (M_f) temperatures define the temperature range within which the transformation takes place.

Martensite has a body-centered tetragonal (BCT) crystal structure, which is distinctly different from the FCC structure of austenite. The transformation from austenite to martensite occurs through a diffusionless shear mechanism [7]. This mechanism involves a highly coordinated movement of atoms, which results in a distortion of the crystal lattice rather than atomic diffusion. This process creates a massive amount of internal strain and introduces a very high density of dislocations in the martensitic structure. These dislocations act as barriers to further dislocation motion, which makes deformation extremely difficult and contributes to the high strength and hardness of martensite.

In low-carbon steels, martensite typically forms as lath martensite, with fine, parallel plates (laths) oriented in the same direction, creating a needle-like structure. In high-carbon steels, martensite forms as finer, elongated needle-like structures. These needles can be extremely small (microscopic in size), and their high dislocation density contributes to the steel's hardness. As carbon content increases, the size of the laths diminishes (Figure 2.7).

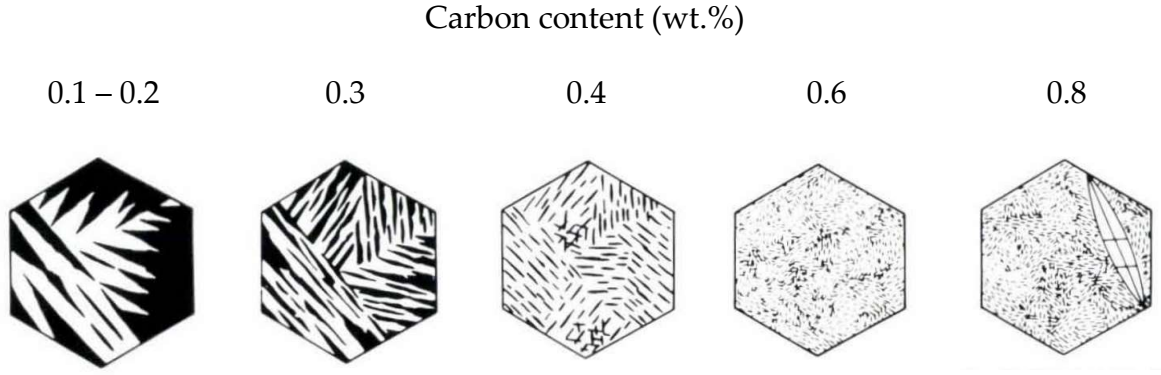


Figure 2.7. Schematic illustrations depicting morphological changes in lath martensite structures with varying carbon content in steels [8]

2.1.3. Bainite transformation

Bainite transformation is the intermediate phase between martensite and pearlite. The process is diffusion-controlled regarding carbon partitioning, although it also incorporates shear deformation, rendering it a complex transformation [9]. Bainite develops at temperatures below pearlite transformation and above martensite transformation when austenite is cooled at a rate that is slower than that necessary for the creation of martensite, yet faster than that required for pearlite formation.

Bainite consists of a fine mixture of ferrite and cementite. Upper bainite forms at higher temperatures (350–550 °C) and has a feathery morphology, with carbide precipitates between ferrite laths. Lower bainite forms at lower temperatures (250–350 °C) and has a plate-like structure, with fine carbides precipitated within the ferrite plates (Figure 2.8).

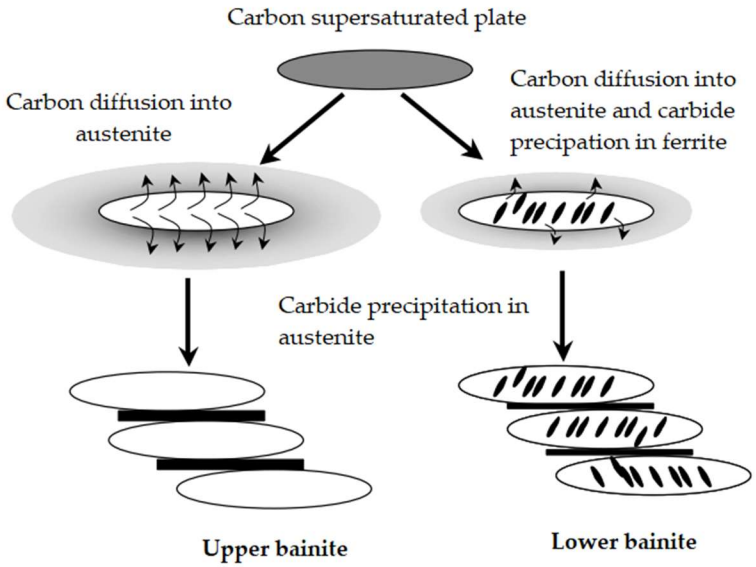


Figure 2.8. Schematic illustration of the bainite transformation in steels [10]

Upper bainite has a slower nucleation rate due to reduced driving force from undercooling, while lower bainite exhibits a higher nucleation rate because of greater undercooling despite lower atomic mobility. In both cases, ferrite grows by a shear mechanism, similar to martensitic transformation, but with carbon partitioning into

the surrounding austenite. In upper bainite, carbides precipitate at the ferrite-austenite interfaces due to higher carbon mobility at elevated temperatures. In contrast, in lower bainite, carbides precipitate within the ferrite plates because of reduced carbon mobility at lower transformation temperatures.

2.2. Hardenability

Hardenability refers to the ability of a ferrous material to harden after austenitization and quenching. This general definition comprises two subdefinitions: the ability to reach a certain hardness level and the uniform hardness distribution within a cross section. The ability to reach a certain hardness level is associated with the highest attainable hardness. It depends first of all on the carbon content in the steel [11]. This doctoral thesis focuses on the latter aspect—the uniform hardness distribution within a cross-section.

High hardenability steels are used for larger machine parts where high hardness is required in deeper layers. Steels with low hardenability may be used for smaller components or for surface hardened components such as gears or shafts.

As mentioned in Section 2.1.2. Martensite transformation, carbon dissolved in the austenite after the austenitizing treatment has relevant influence on the hardness of martensite. The hardness of quenched steel reaches its maximum when the critical cooling rate is achieved during quenching, ensuring that no austenite remains. However, the carbon content (wt.%) in steel imposes a fundamental limit on hardness, even when the microstructure consists entirely of martensite (Figure 2.9). Additionally, the proportion of martensite in the microstructure plays a crucial role in determining the final hardness after quenching.

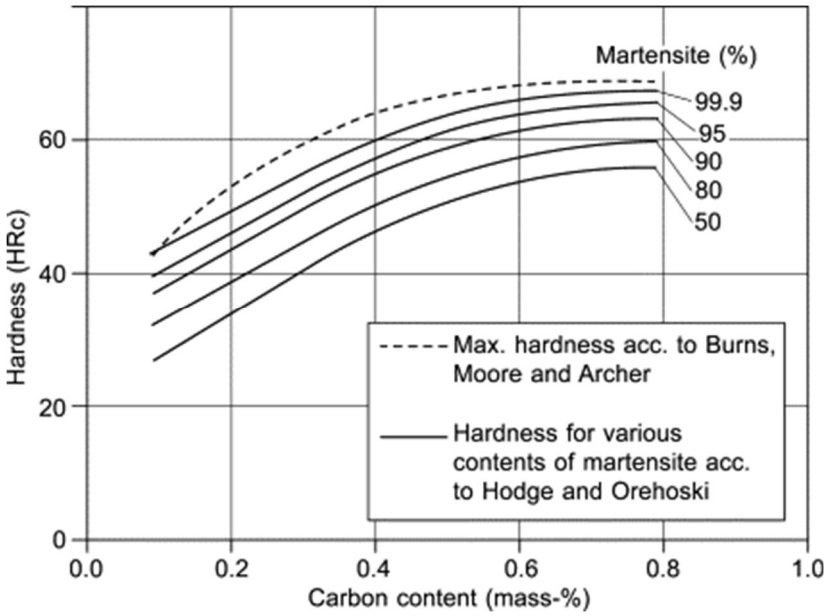


Figure 2.9. Achievable maximum hardness influenced by the carbon content and amount of the martensite in the microstructure [12].

Increased carbon concentration in steel stabilizes austenite and delays its transformation into ferrite. Higher carbon content lowers the equilibrium temperature for the austenite-to-ferrite (A_3) transformation. The transformation of austenite to ferrite requires the diffusion of carbon out of the austenite lattice, as ferrite has an extremely low solubility for carbon. A higher carbon concentration increases the amount of carbon that must diffuse, thereby slowing the transformation process.

Carbon increases the lattice distortion of the BCT structure of martensite. The higher the carbon content, the greater the distortion, which directly enhances hardness. While the carbon content significantly affects the maximum hardness achievable, it has little impact on the formation of martensite in the interior of the material.

Upon cooling, the alloying elements, as well as carbon, are reallocated between ferrite and cementite. The redistribution of alloying elements requires diffusion. The diffusivity of alloying elements is significantly lower than that of carbon in steel. As a result, austenite is difficult to transform to pearlite when alloying elements are present. The critical cooling rate is reduced by alloying elements in steel. Thus, the

alloying elements improve hardenability [13, 14]. Alloying elements will effectively enhance deep hardenability only if they form a complete homogeneous solution with austenite.

2.2.1. Jominy test

Widely used experimental method for characterizing the hardenability was proposed by Jominy and Boegehold in 1938 [15]. A standardized bar, 25.4 mm diameter (1 inch) and 101.6 mm long (4 inch), is heated to the austenitizing temperature and then placed on a rig in which one end of the rod is cooled by a standard jet of water (Figure 2.10). This leads to a gradual reduction in the cooling rate along the bar, from rapid cooling at the quenched end to rates comparable to air cooling at the opposite end.

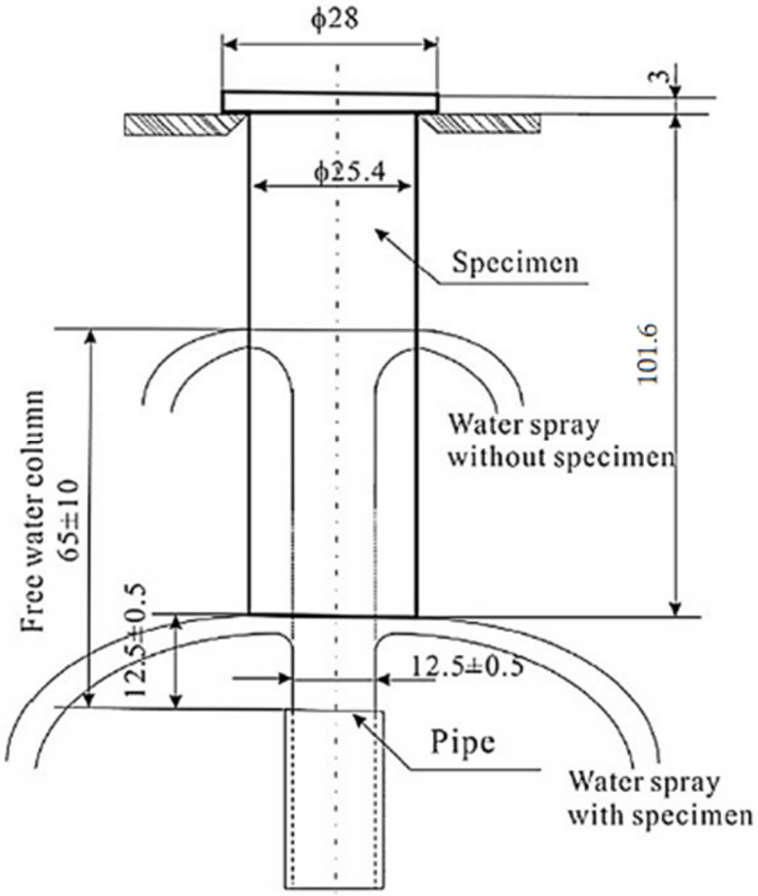


Figure 2.10. Schematic representation of the Jominy- Boegehold test [16]

The round specimen is then ground flat along its length to a depth of 0.38 mm (15 thousandths of an inch) to remove decarburized material. The hardness is measured at specific distances from the quenched end of Jominy specimen: 1.5, 3, 5, 7, 9, 11, 13, 15, 20, 25, 30 40, 50 mm respectively. Different cooling rates result in variations in the final microstructure and, consequently, in different hardness (Figure 2.11). Higher hardness occurs where higher volume fractions of martensite develop. Lower hardness indicates transformation to bainite or ferrite/pearlite microstructures. The distance from the quenched end, over which martensite is obtained, is then the measure of hardenability (up to 50% of martensite in the microstructure).

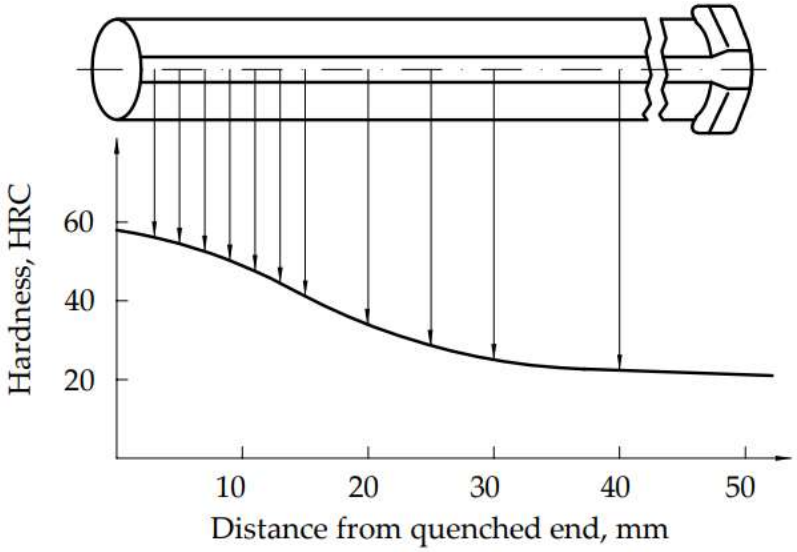


Figure 2.11. Measuring hardness and cooling rates on the standard Jominy specimen and resulting Jominy curve

For two distinct Jominy curves it is visible that a different hardness value might be attained with the same microstructure (i.e. 99% martensite) at the same distance (i.e. 5 mm) from the quenched end (Figure 2.12).

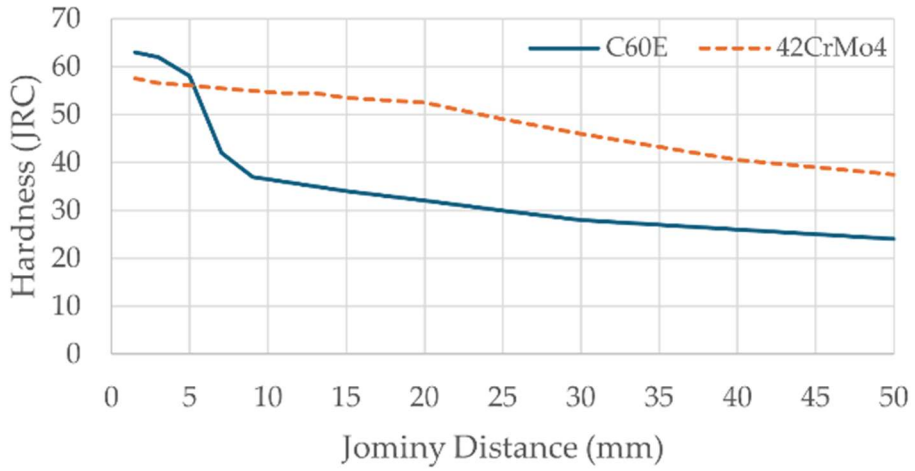


Figure 2.12. Jominy curves for steels C60E and 42CrMo4

The variation is attributed to the differences in chemical composition, primarily the carbon content. Conversely, it is possible to obtain the same hardness value at an equivalent distance, yet with a different microstructure. It originates from the differing chemical composition (Figure 2.13).

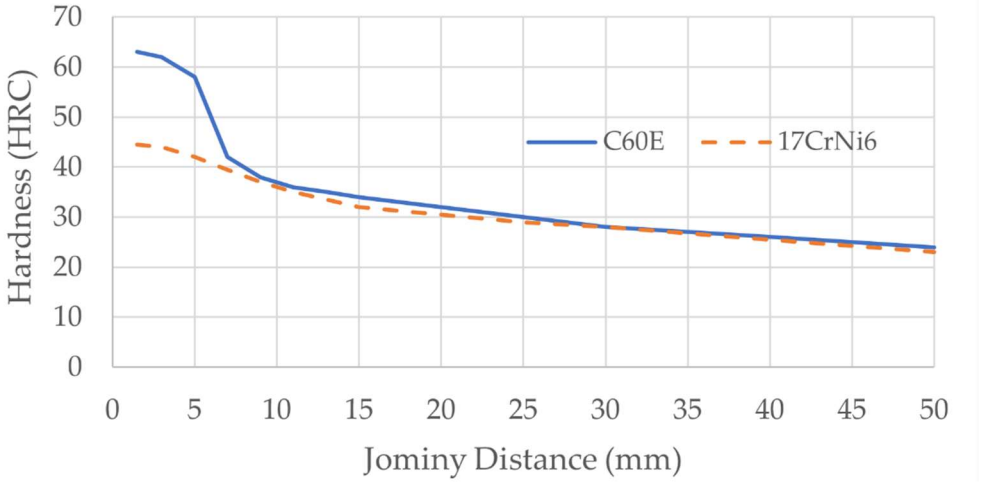


Figure 2.13. Jominy curves for steels C60E and 17CrNi6

The cooling process in the Jominy test occurs primarily through heat conduction, influenced by physical properties such as density, specific heat capacity (c), thermal conductivity (λ), and heat transfer coefficient (α). These properties remain mostly

unaffected by the chemical composition of the steel. While the cooling rate is the same at specific distances from the quenched front, the chemical composition of the steel affects the resulting microstructure and phase transformations at these distances, as different alloying elements influence the transformation kinetics and the formation of phases such as martensite, bainite or pearlite.

2.2.2. Influence of alloying elements on hardenability

Steels with different chemical compositions exhibit distinct Jominy hardness profiles, leading to differences in hardenability [17]. The impact of alloying elements is observed in the delayed pearlite transformation [18]. Thus, the Ms temperature, necessary for martensitic transformation, is reached even at lower cooling rates.

The martensitic structure exhibits high hardness; therefore, steels with high hardenability will maintain elevated hardness levels even at greater distances from the quenched end. As stated in the chapter's introduction, the maximum achievable hardness of steel is largely determined by its carbon content. As a result, steels with various carbon contents have different hardness, even if they acquire a practically similar martensitic structure at a smaller distance from the quenched end. Steels with high hardenability exhibit a more gradual decline in the Jominy curve. Figure 2.12 shows that 42CrMo4 is a high-hardenability steel, whereas C60E is a low-hardenability steel, despite C60E exhibiting higher hardness at shorter distances from the quenched end.

Carbon increases the stability of austenite due to its high solubility. Carbon lowers the temperature at which this transformation occurs, making martensite formation easier at lower temperatures. The maximal achievable hardness primarily depends on carbon content because carbon strengthens the martensitic lattice by distorting it through interstitial solid solution strengthening.

Chromium is a strong carbide-forming element that produces complex carbides such as Cr_7C_3 and Cr_3C_2. These carbides dissolve gradually in austenite at high temperatures, helping to resist grain coarsening during heat treatment. Chromium enhances steel hardenability by lowering the critical cooling rate, allowing for deeper martensite formation during quenching [19].

With sufficient austenitizing time, chromium significantly increases the depth of hardening by stabilizing austenite at high temperatures, facilitating martensite formation even in thicker sections and under slower cooling conditions. This makes chromium essential in steels requiring high strength and wear resistance, such as tool steels, high-strength structural steels, and automotive components.

Additionally, chromium carbides influence the distribution of carbon in the austenite matrix, indirectly affecting martensitic transformation. By tying up carbon in localized regions, these carbides reduce the carbon content of the surrounding austenite, which raises the martensite start (M_s) temperature and can influence the final hardness of the martensite.

Manganese, as a carbide former (primarily in the form of Mn_3C), plays a crucial role in increasing the hardenability of steel. By forming carbide precipitates, it enhances the steel's ability to form martensite during quenching. In structural steels, manganese raises the critical cooling rate required to suppress pearlite and ferrite formation, enabling deeper hardening and ensuring uniform hardness, even in thicker sections.

Additionally, manganese increases carbon solubility in austenite by expanding the interstitial lattice spaces, allowing more carbon to dissolve. This effect lowers the martensite start (M_s) temperature, extending the time available for transformation during cooling and facilitating martensite formation at slower cooling rates or in deeper layers.

Molybdenum, depending on its composition and carbon content, can form complex carbides such as Mo_2C, which enhance hardness and wear resistance. It significantly

improves hardenability, especially when combined with other elements like nickel and chromium. Molybdenum reduces the rate of phase transformations (e.g., austenite to ferrite), lowering the critical cooling rate and increasing hardening depth, making it crucial for steels that require deep hardening and high strength.

Molybdenum's large atomic size slows diffusion in the steel matrix, decelerating transformations like austenite to ferrite or pearlite, while promoting martensite formation during cooling. Additionally, it forms stable carbides at elevated temperatures, depleting carbon in specific regions. This carbon depletion hinders diffusion-driven transformations, such as pearlite or bainite formation. Molybdenum and chromium carbides precipitate at grain boundaries or within grains, refining the microstructure and reducing the size of regions where diffusional transformations can nucleate.

Nickel enhances hardenability by reducing the critical cooling rate required for quenching and lowers the A_c and A_r critical temperatures, which influences the austenite-to-martensite transition. Unlike other alloying elements, nickel does not form carbides. Alloy steels containing nickel and chromium exhibit higher elastic ratios, increased hardenability, and improved impact strength and fatigue resistance compared to carbon steels.

Silicon slows the pearlite transformation by inhibiting the nucleation and growth of cementite. As silicon is insoluble in cementite, it increases the free energy of the cementite phase, making its formation less favorable. In hypoeutectoid steels, silicon promotes ferrite formation, shifting the transformation toward a ferrite-dominated microstructure. Additionally, silicon enhances the stability of retained austenite by suppressing carbide precipitation during quenching and tempering.

Though it does not form carbides, silicon elevates critical temperatures in proportion to carbon content, requiring higher austenitizing temperatures. While offering a

modest increase in hardenability, silicon is primarily used in small amounts as an effective deoxidizer [20].

Chemical elements, beyond their distinct impact on hardenability, also exhibit a synergistic effect [21]. By carefully selecting and combining elements such as carbon, chromium, manganese, molybdenum, and nickel, engineers can optimize the cooling rate, phase transformations, and final microstructure to meet the specific requirements of various applications, particularly for large or thick sections where precise hardenability is essential.

2.3. Deep Learning

Artificial Intelligence (AI) has become one of the most revolutionary technologies of the 21st century. The rapid progression and extensive applicability of AI have established it as a necessary component of nowadays innovation. The economic implications of AI are huge. Studies suggest that AI could contribute trillions of dollars to the global economy over the coming decade by enhancing productivity and enabling entirely new innovations [22]. The term "Artificial Intelligence" (AI) was first used in a proposal for the Dartmouth summer research project on artificial intelligence in 1955 [23]. Despite its widespread usage today, the term lacks a specific definition. Artificial intelligence (AI) can be defined as the ability of a digital computer or computer-controlled machine to perform tasks commonly associated with intelligent beings, such as learning, comprehension, problem solving and decision-making [24].

In 1959, Arthur L. Samuel revealed methods enabling his checkers program to learn from previously played games. The program improved with each game it played. Samuel named this process as machine learning [25]. Machine learning is considered as a subset of artificial intelligence (AI) that focuses on developing algorithms and statistical models that allow computers to learn from and make predictions or decisions based on data, without being explicitly programmed for every task. Machine

learning models are trained on data to detect patterns and generate conclusions or predictions. The system's performance progressively becomes better as it is exposed to more data. As a result, trained machine learning models can apply their knowledge to previously untested data [26]. Machine learning relies on the harmonious integration of its fundamental components: input data, features, models, training and testing. Each component contributes to its ability to learn, adapt, and generalize.

The first and most critical element of any machine learning process is the input data. This data serves as the raw material upon which the model is built. The quality and diversity of input data directly impact the performance of the machine learning model.

Once input data is gathered, the next step involves extracting features, which are the relevant attributes or variables from the data. Features act as the lens through which the model interprets the input data. Selecting and engineering appropriate features is a crucial task, as it determines how well the model can learn meaningful patterns.

The mathematical or computational representation of the problem is the core of machine learning. The model functions as a mapping mechanism between the input features and the desired output. Various model types exist, from basic linear regressions to sophisticated neural networks, each specialized for specific uses. The choice of the model depends on the complexity of the problem, the nature of the data, and the desired level of accuracy and interpretability.

The model acquires the ability to identify patterns from the data throughout the training process. The quality of training is influenced by factors like the size of the dataset, the algorithm used, and the duration of the training process. Proper training ensures that the model generalizes well to new data.

Testing evaluates the performance of the trained model on unseen data. This step is vital for assessing how well the model can generalize.

Classical machine learning relies more heavily on human intervention for learning. Human experts identify the features necessary to discern distinctions among data sources, typically necessitating more structured data for effective learning [27].

Deep learning is a form of machine learning that employs neural networks to replicate the learning processes of the human brain [28]. Deep learning represents a contemporary term for a methodology in artificial intelligence known as neural networks. Despite requiring additional data for training, deep learning possesses the capability to adjust to novel situations and correct its own errors.

An artificial neural network consists of up of thousands or even millions of connected basic processing nodes that are simply modeled after the structure of the human brain. The majority of contemporary neural networks are structured in layers of nodes and operate in a "feed-forward" manner, indicating that data travels through them in one direction. A single node may link to multiple nodes in the layer below, from which it receives data, and to multiple nodes in the layer above, to which it passes data.

Each incoming connection to a node in a neural network is assigned a number called a weight. When the network is active, the node receives a different data value — a number — through each of its connections and multiplies each value by the corresponding weight. The node then sums the resulting products to yield a single output. During the training process, all the weights are initially set to random values. The training data is fed into the input layer, where it passes through the network's subsequent layers. In each layer, the data is multiplied by the associated weights, summed, and transformed, until it reaches the output layer. Throughout training, the weights are adjusted iteratively so that the network consistently produces outputs similar to the expected results when the training data with the same labels is provided.

2.3.1. Biological neuron

The structure of the brain's component neurons (nerve cells) must be understood to comprehend the brain's capabilities. The human brain consists of around 100 billion (10^{11}) neurons, organized according to their specific functions in a meticulously determined manner. Each neuron may be connected up to 10.000 other neurons. A neuron comprises four fundamental components: the cell body (soma), a collection of dendrites (branches), an axon (elongated structures that convey electrical signals), and a series of axon terminals [29]. The structure of biological neuron is illustrated in Figure 2.14.

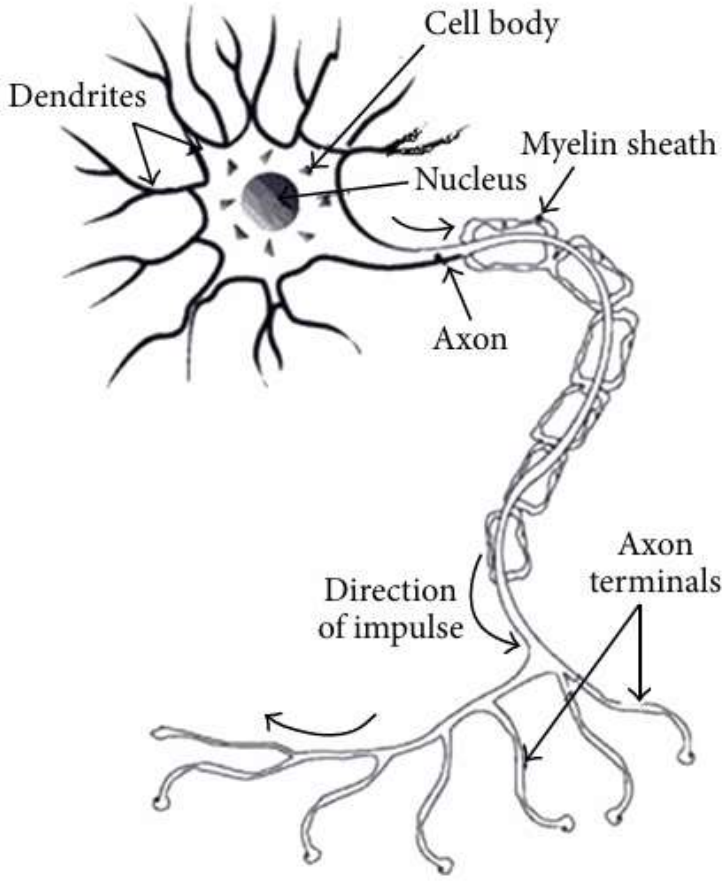


Figure 2.14. Structure of biological neuron [30]

2.3.2. Processing Element

In 1943, neuroscientist Warren McCulloch and logician Walter Pitts developed a model that provided one of the earliest mathematical abstractions of a biological neuron [31].

McCulloch and Pitts outlined five key assumptions governing the operation of neurons. First, the neuron can exist in one of two possible states: active (1) or inactive (0), following an "all-or-none" principle. Second, each neuron has a specific predefined threshold, meaning that a predetermined number of synapses must be activated to stimulate the neuron at any given moment. Third, the neuron receives inputs from excitatory synapses, each with a uniform weight. It also receives inputs from inhibitory synapses, which have a definitive effect — if the inhibitory synapse is activated, the neuron cannot fire. Fourth, a time quantum for the integration of synaptic inputs is based on the physiologically observable synaptic delay. The fifth assumption states that the structure of the neuron remains constant over time.

In the quantum interval, the neuron reacts to the activity of its synapses. The neuron sums its synaptic inputs and determines whether the total reaches or surpasses its predefined threshold if no inhibitory synapses are engaged. Upon reaching or exceeding its threshold, the neuron subsequently activates. If it fails to do so, the neuron remains inactive. Mathematically. their model treated the neuron as a binary threshold unit, meaning: Inputs are either 0 (inactive) or 1 (active) and each input has a weight and. If the weighted sum of inputs exceeds a threshold, the neuron fires (outputs 1); otherwise, it stays inactive (outputs 0):

$$y = \begin{cases} 1 \text{ if } \sum_{i=1}^{n} x_i w_i \geq \Theta \\ 0 \quad \text{otherwise} \end{cases} \tag{2.3}$$

Where x_i are the binary inputs (0 or 1), w_i are the corresponding weights, Θ is the threshold and y is the output (1 if activated, 0 if not).

Biological neurons do not operate as mere binary switches. They function by a complex combination of electrical and chemical processes that allow them to interpret information in extremely advanced manners [32, 33].

McCulloch-Pitts neuron was designed to simulate logical operations, and it is highly simplified biological neuron. It is deterministic and do not include a learning mechanism. Despite its limitations (absence of a learning mechanism, fixed weights assigned to all inputs, and inability to handle non-Boolean inputs), the McCulloch-Pitts model is considered as foundational to contemporary cognitive science and neuroscience, especially in the field of artificial intelligence [34].

In 1957, psychologist Frank Rosenblatt developed an adaptable system known as the perceptron [35, 36]. Perceptrons are highly simplified models of biological brains, created to explore the mechanisms by which the brain performs its functions. Rosenblatt designed his perceptron based on the structure and function of the optic nerve [37, 38]. Specifically, he used the retina as a model, as its primary function is to capture light stimuli. The retina then transforms these stimuli through a complex chemical process into electrical impulses, which are transmitted to the brain [39].

Rosenblatt's perceptron comprises three components (layers): sensory unit (S-Unit), associator unit (A-Unit), and response unit (R-Unit). Rosenblatt also took into consideration four-layer systems (schematic type S—A^{I}—A^{II} -R), which have two layers of associator units [40, 41]. Each layer comprises a set of simple threshold elements ('neurons') [42, 43]. Figure 2.15 depicts a simple perceptron.

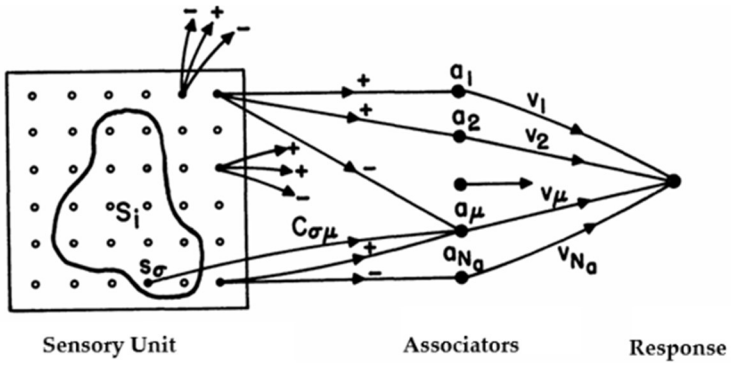


Figure 2.15. Rosenblatt's simple perceptron [42]

From the perspective of modern neural network terminology, Rosenblatt's perceptron is typically described as a single-layer network. This refers to the fact that there is only one layer of adjustable weights.

Rosenblatt's crucial contribution was the development of the perceptron learning rule, allowing the system to learn patterns from data rather than being a static logical unit such as MP neuron. The notion that learning takes place through weight adjustments based on input-output interactions was inspired by the Hebbian rule [44].

Only the connections from A-units to R-units are determined by synaptic plasticity and therefore possess adjustable weights. [43, 45].

To simplify computation, in his initial perceptron model, Rosenblatt introduced the bias term (b), expressed as the negative equivalent of the threshold (Θ):

$$b = -\Theta \tag{2.4}.$$

In that way, the threshold was incorporated as part of the perceptron's decision rule and suggested that it could be treated as an additional input with a constant input $x_0 = 1$, with a weight $w_0 = -\Theta$. This reformulated a linear function:

$$y = \mathbf{w} \cdot \mathbf{x} + \mathrm{b}$$
$$y = \sum_{i=1}^{n} x_{\mathrm{i}} w_{\mathrm{i}} - \Theta \tag{2.5},$$

into:

$$y = \sum_{i=0}^{n} x_i w_i \tag{2.6}.$$

After computing the weighted sum, the perceptron applies a (sign) step function:

$$o = \mathrm{sign}(y) = \begin{cases} +1 & \text{if } y > 0 \\ -1 & \text{otherwise} \end{cases} \tag{2.7}.$$

According to the Perceptron Convergence Theorem, the perceptron learning algorithm is guaranteed to converge only if the dataset is linearly separable and a finite learning rate is used. When these conditions are met, the perceptron can find a set of weights that separates the classes.

The perceptron learning rule adjusts weights proportionally to the product of the input (x_i) and the error signal ($t-o$), modifying weights only when there is a misalignment between the target (t) and the output (o). The perceptron learning rule can be explained by the expression:

$$w_i \leftarrow w_i + \eta(t - o)x_i \quad (2.8),$$

where w_i is weight for the i-th input, x_i is i-th input, t is true output (target), o is predicted output, η is a learning rate.

The perceptron exhibited clear limitations. The perceptron algorithm could only solve problems where a single linear decision boundary can perfectly separate the classes. In a case when data is not linearly separable (e.g., XOR problem), the perceptron fails to find a weight vector that minimizes the classification error. The perceptron has no mechanism to directly measure or minimize classification error, as it does not optimize a specific cost function to assess overall performance. It simply updates weights whenever it encounters a misclassification. As a result, the perceptron keeps oscillating through the dataset without settling on a stable solution.

The use of a step function introduces non-linearity in the output (e.g., +1+ or −1-), but this does not address the issue of non-linear separability in the data.

The key contribution of Rosenblatt's perceptron is its introduction of a learning rule for adjusting weights based on errors.

Bernard Widrow and Marcian E. Hoff created ADALINE (ADAptive LInear NEeuron), an early single-layer artificial neural network, shortly after Rosenblatt's

perceptron. They also constructed a compact electrical device that could be trained, serving as a hardware manifestation of the algorithm they developed [46].

It pertained to perceptrons. It was referred to as an adaptive neuron with variable connection strengths (weights). The adaptive neuron calculated a weighted sum of the input activities (inputs can take values of either a +1 or a -1) multiplied by the synaptic weights, in addition to a bias element. The output became + 1 if the sum exceeded zero. If it was less than or equal to zero, the output was -1.

The ADALINE works by minimizing the mean square error of the predictions generated by a linear function. This indicates that the learning process is not dependent on the results of a threshold function as in the perceptron, but rather on the results of a linear function. The developed algorithm is known as Least Mean Squares (LMS) algorithm. The algorithm adjusts the weights in the ADALINE model incrementally during training, using a gradient descent optimization approach.

The algorithm initiates by multiplying all input values ($\mathbf{x_i}$) by their respective weights ($\mathbf{w}$), following summing these products to provide the net input (y), including of bias, equivalent to Rosenblatt's perceptron. In vector form, it is represented by an expression:

$$y_i = \mathbf{x}_\mathrm{i}^\mathrm{T}\mathbf{w} = \mathbf{w}^\mathrm{T}\mathbf{x_i} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \mathbf{w_j}\,\mathbf{x_{ij}} \qquad (2.9),$$

where i is i-th pattern (sample of inputs or observation), j is j-th number of observations, $1 \leq j \leq n$.

The error is calculated as the difference between the calculated (y_i) and desired output (d_i):

$$e_\mathrm{i} = d_\mathrm{i} - y_\mathrm{i} = d_\mathrm{i} - \mathbf{x}_i^\mathrm{T}\mathbf{w} \qquad (2.10).$$

The objective is minimizing the mean square of the error over the set of training patterns. To eliminate the influence of negative values while calculating the mean, the error is squared:

$$e_{\mathrm{i}}^2 = (d_{\mathrm{i}} - y_{\mathrm{i}})^2 = d_{\mathrm{i}}^2 - 2d_{\mathrm{i}}\mathbf{x}_{\mathrm{i}}^{\mathrm{T}}\mathbf{w} + \mathbf{w}^{\mathrm{T}}\mathbf{x_i}\,\mathbf{x}_{\mathrm{i}}^{\mathrm{T}}\mathbf{w} \qquad (2.11).$$

The mean square error (expected value) is represented as:

$$\mathrm{MSE} = \xi = \mathrm{E}(e^2) = \mathrm{E}(d^2) - 2\mathrm{E}(d_{\mathrm{i}}\mathbf{x}_{\mathrm{i}}^{\mathrm{T}})\mathbf{w} + \mathbf{w}^{\mathrm{T}}\mathrm{E}(\mathbf{x_i}\mathbf{x}_{\mathrm{i}}^{\mathrm{T}})\mathbf{w} \qquad (2.12).$$

The weights are fixed during applying the whole set different patterns. The mean square error is quadratic function of the weights. Basically, the goal is to find set of weights, such that MSE is the lowest. To find the best wights ($\mathbf{w}$), the steepest descent (gradient descent) method is used:

$$\mathbf{w}_{\mathrm{j+1}} = \mathbf{w}_{\mathrm{j}} + \mu(-\widehat{\nabla} e_{\mathrm{j}}^2(\mathbf{w})) \qquad (2.13),$$

where $\widehat{\nabla} e_{\mathrm{j}}^2(\mathbf{w}) = \frac{\partial e_{\mathrm{j}}^2(\mathbf{w})}{\partial \mathbf{w}}$ is a gradient estimate, μ is coefficent that control the speed of convergence.

The gradient estimate (instantaneous gradient) approximates the total (full) gradient of the mean squared error, calculated using solely the current training sample [47]. In other word, the LMS algorithm estimates the gradient by using the error from a single training sample.

The gradient estimate is:

$$\widehat{\nabla} e_{\mathrm{j}}^2 = \frac{\partial e_{\mathrm{j}}^2}{\partial \mathbf{w}} = 2e_{\mathrm{j}}\frac{\partial e}{\partial \mathbf{w}} = 2e_{\mathrm{j}}(-x_{\mathrm{j}}) \qquad (2.14).$$

The outcome of equation (2.13) can be integrated into equation (2.12), yielding the LMS algorithm:

$$\mathbf{w}_{\mathrm{j+1}} = \mathbf{w}_{\mathrm{j}} + 2\mu e_{\mathrm{j}}\mathbf{x_j} \qquad (2.15).$$

Nowadays, the LMS algorithm is recognizable in the following form[2]:

$$w_{i+1} \leftarrow w_i + \eta\, e\, x_{\mathrm{i}} \tag{2.16}.$$

While Rosenblatt's perceptron learning rule implicitly reduced error, he did not explicitly define a mathematical function to measure and minimize prediction errors. It was Widrow and Hoff who, with the development of the Least Mean Squares (LMS) algorithm, introduced the concept of using a mathematical function — specifically, the Mean Squared Error (MSE) — to measure and minimize prediction errors in neural network training. Widrow and Hoff formalized the idea of a quantitative measure of error that could guide learning. This was a foundational step because it allowed the learning process to be framed as optimization and it linked the model's performance (error) to its adjustable parameters (weights), creating a systematic way to adjust weights to minimize error.

Before the introduction of optimization frameworks, learning algorithms were often described as rule-based or heuristic-driven processes, with no explicit mathematical structure that guided parameter updates in a way that could be systematically optimized. The LMS algorithm introduced a structured way to approach learning, formalizing it as an optimization problem.

The original Perceptron and Adaptive Linear Neuron were foundational computational units that computed a weighted sum of inputs and then applied a hard threshold function to determine the output. Although these models have their origins in neuroscience and are often referred to as 'neurons', Widrow renamed his ADALINE to Adaptive Linear Element after recognizing that they were computational units with little in common with biological neurons [48].

[2] Sometimes the factor 0.5 is included in the error squared. It is purely for mathematical convenience. It does not affect the outcome of the algorithm, as the learning rate compensates for any constant scaling.

The breakthrough that allowed neural networks to handle nonlinearity came from introducing nonlinear activation functions. This was critical in moving beyond simple linear classifiers and allowing networks to model complex patterns.

Today, the processing element or nodes refers to computational unit. Modern interpretations of processing elements replace the hard threshold with a nonlinear activation function, such as sigmoid, ReLU, or tanh (Figure 2.16).

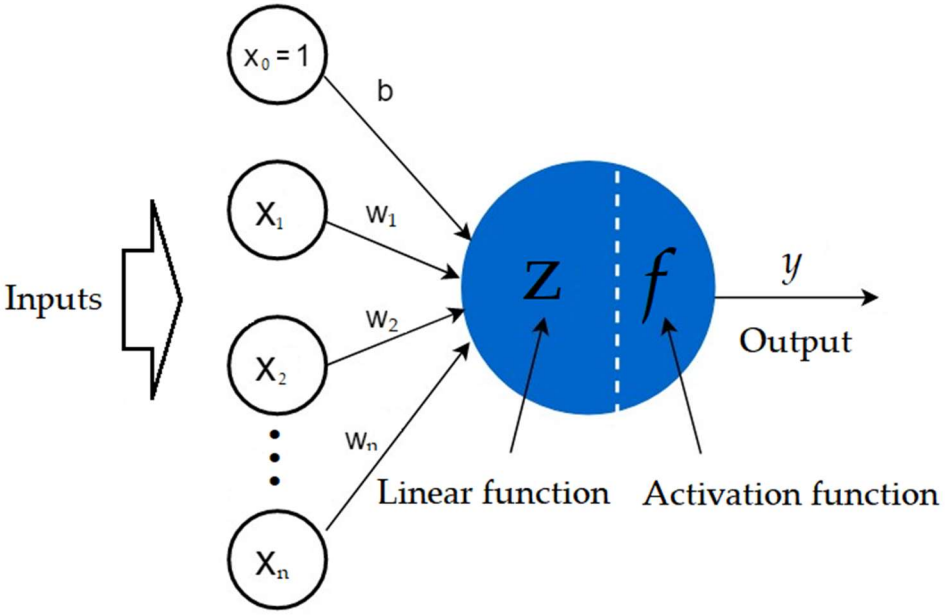


Figure 2.16. Processing element

2.3.3. Multilayered Neural Network

In the nature, truly linear relationships, where changes in one factor lead to proportional changes in another, are rare. Most systems, whether physical, biological, or social, are nonlinear [49]. Material behavior often depends on intricate, multiscale phenomena (e.g., atomic interactions, grain boundaries, phase transformations), which are inherently nonlinear. Nonlinear systems are more challenging to analyze because small changes in input can result in wildly different outputs ('butterfly effect'). Understanding these systems requires tools and methods that can handle their inherent complexity and uncover patterns beyond simple cause-and-effect reasoning.

The transition from single-layer to multilayer neural networks revolutionized artificial intelligence, establishing neural networks as an indispensable tool for solving complex problems of today's computing.

Feedforward Neural Networks (FNNs) are one of the most fundamental and widely used architectures in Artificial Neural Networks (ANNs). In this design, information flows in a unidirectional manner, progressing from the input layer, through one or more hidden layers, to the output layer.

Each layer in the network consists of interconnected neurons (nodes), where every neuron in a layer is connected to all neurons in the subsequent layer. These connections are weighted, with each weight determining the strength and influence of the signal passed between neurons. During the forward pass, input data is multiplied by these weights and processed through an activation function at each neuron, introducing non-linearity to enable the network to model complex relationships (Figure 2.17).

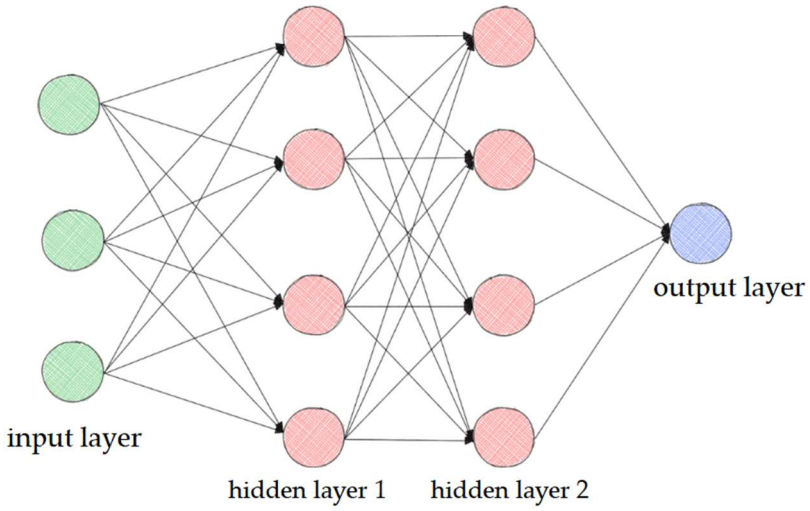


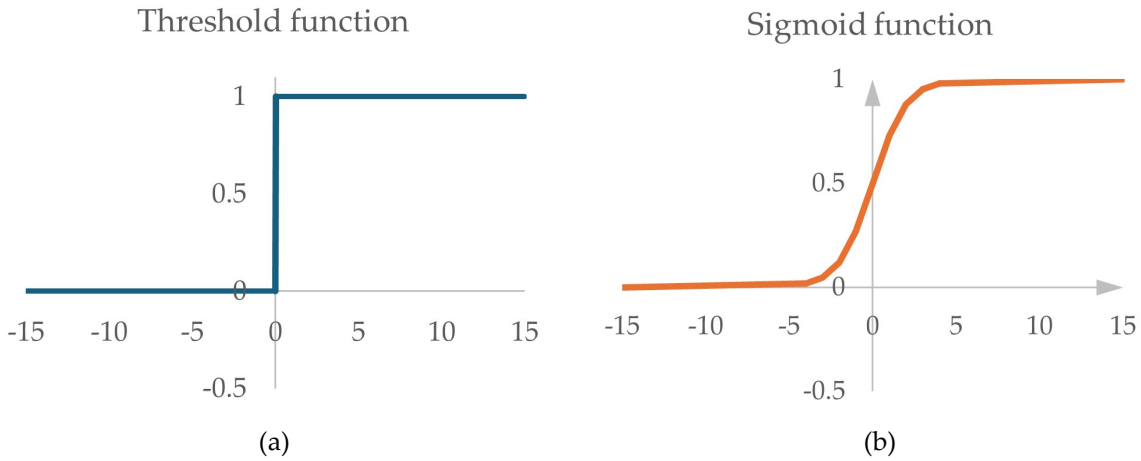
Figure 2.17. Architecture of simple feedforward neural network with two hidden layers

The input layer represents the initial data as a vector of predefined values. The hidden layers, located between the input and output layers, apply non-linear transformations to the data, enabling the network to learn intricate patterns and perform sophisticated

tasks. Finally, the output layer processes the data from the last hidden layer to produce the final results. This fully interconnected structure allows FNNs to efficiently capture relationships between inputs and outputs, making them a powerful tool for various applications in machine learning [50].

In a single-layer networks, learning algorithms could be applied. Multilayer networks introduced hidden layers, which made it unclear how to propagate the error signal back through the layers to update weights. Backpropagation solved this problem by providing a systematic way to compute gradients for all weights in the network [51].

Backpropagation Algorithm, popularized by Rumelhart, Hinton, and Williams, allowed the efficient training of multilayer networks using gradient descent to minimize error [52]. This error is typically measured by a cost function, such as the Mean Squared Error (MSE), which quantifies the difference between the predicted output and the true target values. The introduction of hidden layers including nonlinear activation functions in these models was groundbreaking. "Semilinear" is the term they used to characterize the required function with the desired activation function features. They underscored the necessity for differentiability and nonlinearity. The word "semilinear" is employed to characterize the sigmoid function's "soft thresholding" characteristic, which integrates linearity in localized regions with overall nonlinearity (Figure 2.18).



(a) (b)

Figure 2.18. Threshold (a) and sigmoid function (b)

In contrast to the step function, the sigmoid function is differentiable, a crucial attribute for employing gradient descent and backpropagation in the training of neural networks.

The sigmoid function is denoted by the formula:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.17}$$

A very useful property of the sigmoid function is that the derivative of the sigmoid function can be expressed in terms of the sigmoid function itself:

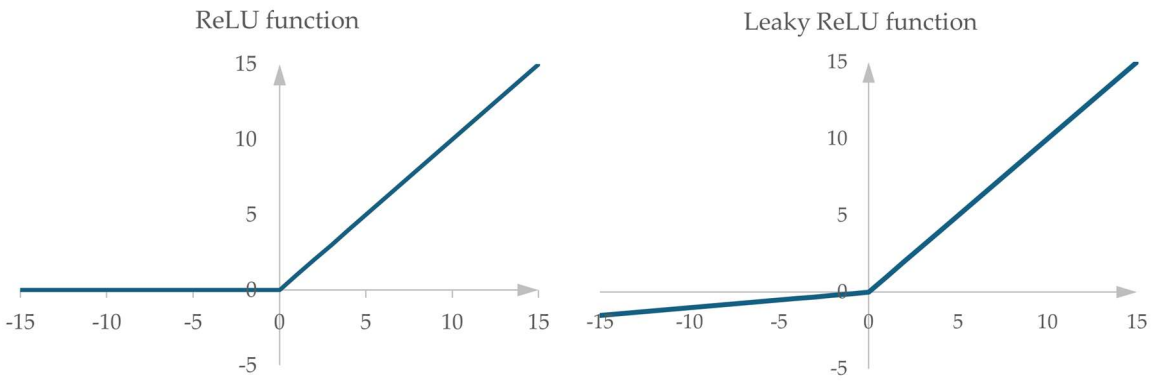
$$\frac{d\sigma(x)}{dx} = \sigma(x) \cdot (1 - \sigma(x)) \tag{2.18}$$

This compact form enhances the efficiency of backpropagation computations in neural networks. But, when the value of sigmoid function is near 0 or 1, the derivative becomes very small, which can lead to the vanishing gradient problem. This is one reason why some other activation functions are used in modern architectures (beside sigmoid). The ReLU (Rectified Linear Unit) activation function was popularized by Vinod Nair and Geoffrey Hinton [53]. The ReLU activation function (which outputs zero for negative inputs and the input itself for positive inputs) has become widely popular because for $x > 0$, the gradient is constant (equal to 1). This avoids the vanishing gradient problem in the positive activation regions. If a neuron frequently outputs negative values during training, its weights may stop updating altogether because the gradient becomes zero**.**

Leaky ReLU function allows a small, non-zero gradient for $x \leq 0$, and it is defined as:

$$f(x) = \begin{cases} x & \text{if } x > 0, \\ \alpha x & \text{if } x \leq 0, \ \alpha > 0 \end{cases} \tag{2.19}$$

This guarantees that neurons can recuperate even from negative weighted sum (input) and prevents the gradient from becoming exactly zero. The ReLU function and Leaky ReLU function are illustrated in Figure 2.19



(a) (b)

Figure 2.19. ReLU (a) and Leaky ReLU (b) function

The activation a_j^l of the j^{th} neuron in the l^{th} layer is related to the activations in the $(l-1)^{th}$ layer by the equation:

$$a_j^l = z\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right) \tag{2.20}.$$

The activation a_j^l of the j^{th} neuron in the l^{th} layer is schematically represented in the Figure 2.20.

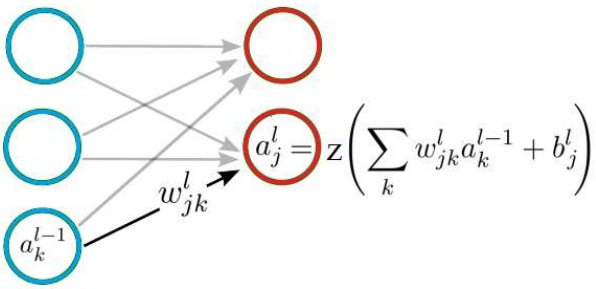


Figure 2.20. Schematic representation of the two neurons in the hidden layer *l*, fully coupled to neurons from the preceding hidden layer (*l*-1)

The cost function in a neural network is not always the mean squared error (MSE), but MSE is one of the most commonly used cost functions, especially in regression problems. Mean Squared Error (MSE) is expressed as:

$$C = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2.21),$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, and n is the number of examples.

MSE is often utilized in tasks involving regression, where the target variable is continuous, meaning it can take any real number. The cost function in the backpropagation algorithm can be computed either for individual training examples or for all training examples, depending on how the training process is structured.

When training a neural network using stochastic gradient descent (SGD) or its mini-batch variant, the cost function is typically computed for a single training example or a small batch of examples. The cost for a single training example x is:

$$C(x) = \frac{1}{2}(y - \hat{y})^2 \quad (2.22).$$

When considering the entire dataset, the overall cost function (often referred as global cost function) is computed as the average (or sum) of the costs for all training examples. For n training examples, the cost is:

$$C = \frac{1}{n}\sum_{i=1}^{n} C(x_i) \quad (2.23).$$

The learning process uses the chain rule to compute gradients layer by layer. Hidden units adjust their activations indirectly by altering weights and biases to enhance their contribution to the desired input-output mapping.

2.4. **Modelling of hardenability**

Chemical composition of steel, as mentioned in Section 2.2, has significant influence on the hardenability. Many efforts have been undertaken to define hardenability according to its chemical composition. A precise model for calculating hardenability

(Jominy curve) at the initial phase of steel production could enable control over the hardenability of the end product. Grossmann characterized hardenability by determining the ideal critical diameter as the maximum diameter of a cylindrical specimen that converts into a minimum of 50% martensite when subjected to quenching at an endlessly large cooling rate at the surface [54]. In his subsequent research, Grossmann suggested the use of multiplicative coefficients for alloying elements to compute the critical diameter [55].

Crafts and Lamont explored the factors influencing the hardenability of steel, particularly the role of alloying elements and cooling rates (quenchants). They emphasized the Jominy end-quench test as a key method for measuring hardenability [56]. Kramer et al. systematically examined how different alloying elements influence steel hardenability. Their research evaluated the impact of sixteen alloying elements using. They demonstrated that certain elements, especially molybdenum and chromium, significantly improve hardenability by delaying pearlite and bainite formation, allowing for deeper martensitic hardening [57]. Kramer et. al. provided a structured methodology to understand and predict how various alloying elements affect steel's hardenability. Comstock investigated how titanium affects the hardenability of steel. The study highlights that titanium has a complex effect—it can both increase and decrease hardenability depending on its interaction with carbon and nitrogen [58]. Hodge and Orehoski examined how carbon content influences hardenability in low-alloy steels. Their research showed that as carbon content increases, hardenability improves, primarily due to carbon's role in stabilizing austenite and promoting martensite formation during quenching [59]. They also proposed formula to estimate the hardness of steel with approximately 50% martensite using multiplicative coefficients for various alloying elements. Just introduced innovative formulas aimed at predicting the hardenability of steels based on their chemical composition. His work focused on correlating the average

compositions of steels with their Jominy end-quench hardness profiles, providing a more precise method for estimating hardenability curves [60].

Brown and James focused on improving the precision of hardenability estimation in steels. They discussed advancements in experimental measurement techniques, mathematical modeling, and process control methods to predict hardenability more accurately [61].

Kunze and Russel, as well as Doane [19] proposed improved empirical and statistical methods to predict hardenability [62, 63]. Mangonon investigated how molybdenum (Mo) and vanadium (V) influence the hardenability of 4330 alloy steel. His research highlighted that molybdenum significantly enhances hardenability by delaying pearlite and bainite formation, while vanadium refines grain size and contributes to precipitation strengthening. He indicated that their combined effect is not purely additive, as interactions between molybdenum and vanadium influence carbide formation and austenite stability [64]. Tartaglia et. al. determined the effects of manganese, nickel, chromium and molybdenum on annealability of carburizing steels. They showed that substitution of the ferrite-stabilizing elements such as chromium and molybdenum with the austenite-stabilizing elements such as manganese and nickel is generally an effective means of reducing the annealing time while maintaining constant hardenability [65]. Kasuya and Yurioka introduced an improved carbon equivalent (CE) formula and multiplying factors to more accurately predict the hardenability of steel. They incorporated the effects of alloying elements like Mn, Cr, Mo, and B, providing a more precise estimation of steel's reaction to heat treatment [66]**.** Yamada et al. examined how chromium (Cr) and other alloying elements influence the different steel properties. Their research discovered that increasing chromium content (1–3%) enhances hardenability and strength but can negatively impact toughness and stress corrosion cracking (SCC) resistance, depending on the presence of additional elements like molybdenum and nickel. [67].

The correlation between the values derived from the Jominy end-quench test and the chemical composition of steel, and vice versa, cannot be accurately established by any mathematical function. Advanced regression analysis is required. At the transition between the 20th and 21st centuries, developments in computers and software have positioned artificial neural networks as important computational methods in materials science. Deep learning, as described in Section 2.3. is a powerful tool for finding patterns in multi-dimensional data. Deep learning uses algorithms such a computer can learn from empirical dataset by modelling nonlinear relationships between the material properties and influencing factors. Artificial neural networks (ANNs) are widely used in modeling steel and metal alloy issue due to their efficiency in handling regression tasks. ANNs are characterized by their ability to learn from labeled dataset and are, therefore, well suited for supervised learning applications.

Bhadeshia analyzed the application of artificial neural networks (ANNs) in the field of materials science, focusing on their ability to model complex relationships within materials behavior, particularly in steel processing. His research provided an overview of how neural networks can be used to predict and optimize various material properties, such as hardenability, microstructure formation, phase transformations, and material performance under different conditions [68].

Filetin et al. demonstrated how neural networks can be used to predict steel properties, particularly when some of the influencing factors and their relationships are not known. They proposed a multi-layer feedforward neural network to predict the Jominy curve based on chemical composition [69]. Studies of Sitek et. al. highlighted the growing influence of artificial intelligence and machine learning techniques (specifically neural networks) in the material design and processing of steels. They demonstrate that neural networks can effectively model and predict the properties of steels, ranging from constructional steels to high-speed steels, by analyzing the chemical composition and processing conditions [70-75].

Vanucci and Colla proprosed an approach to design the chemical composition of steel, based on neural networks and genetic algorithms, which aims to achieve a desired hardenability behavior [76]. Geng X. et al established a combined machine learning (CML) model including k-nearest neighbor and random forest to predict the hardenability curves of non-boron steels based on chemical composition [77].

For non-boron steels, ASTM A255 – 10 provides a method to estimate the ideal diameter (DI) using an empirical formula based on chemical composition, where DI represents the diameter of a round steel bar that forms 50% martensite at its center under ideal quenching conditions. This formula, derived from experimental data, is widely used to predict the hardenability of low-alloy steels [78].

Despite the success of numerous researchers in hardenability modeling, challenges remain, including the need for larger and more diverse datasets, improved model interpretability, and better transferability to new domains. Addressing these issues will further establish neural networks as indispensable tools in materials science. The creation of a representative dataset is crucial for developing an effective model based on artificial neural networks.

3. DATA PREPARATION

Data preparation is a crucial step in developing model by using Artificial Neural Network (ANN). Properly prepared data ensures that the model trains effectively and produces accurate predictions. In general, there are several phases that one has to make during data preparation. The first one is the most obvious one and that is collection of data. Data collecting must fulfill two critical criteria: the data must originate from credible sources, such as experiments, simulations, or databases, and a sufficiently large and representative dataset relevant to the topic at hand must be supplied.

3.1. Data Collection

Overall, dataset for 470 steels were collected. Dataset used for neural network training are obtained from the former Stalowa Wola Steelworks, Poland (about 80%), the standards, steel producer's catalogues, trade literature and Max Planck atlas for heat treatment of steel. The steels investigated are classified according to the following standards EN 10083-2 (Steels for quenching and tempering – Part 2: Technical delivery conditions for non alloy steels), EN 10083-3 (Steels for quenching and tempering – Part 3: Technical delivery conditions for alloy steels), EN 10084 (Case hardening steels – Technical delivery conditions) and similar. Full list of steel grades is given in Table 3.1. While the study focused on 30 steel grades, multiple chemical compositions were examined within each grade. As a result, a total of 470 distinct steel compositions were analyzed.

Table 3.1. Complete list of investigated steel grades

Steel	**Standard**
C35E	EN 10083 – 2 (Steels for quenching and tempering – Part 2: Technical delivery conditions for non alloy steels)
C45E	
C55E	
C60E	
34CrNiMo6	EN 10083 – 3 (Steels for quenching and tempering – Part 3: Technical delivery conditions for non alloy steels)
39NiCrMo3	
25CrMo4	
34CrMo4	
42CrMo4	
50CrMo4	
34Cr4	
38Cr2	
46Cr2	
37Cr4	
41Cr4	
20MnCr5	EN 10084 (Case hardening steels – Technical delivery conditions)
16MnCr5	
20MoCrS4	
17NiCrMo6-4	
20NiCrMo2-2	
18CrNiMo7-6	
16MnCr5	
16MnCrS5	
20MnCr5	
20MnCrS5	
18CrMo4	
55Cr3	Not specified*
15CrMo5	Not specified*
20CrMo5	Not specified*
20MnCr6	Not specified*

* Recognized steel grade, but not listed in a formal standard

The input data consists of 13 hardness measurements obtained at various distances from the quenched end of the Jominy specimen, specifically at 1.5, 3, 5, 7, 9, 11, 13, 15, 20, 25, 30, 40 and 50 mm. Hardness values are reported on the Rockwell C scale (HRC). Additionally, 13 values representing the presence of martensite in the steel's microstructure are included as part of the input data. The output data includes the weight percentages of seven key alloying elements: carbon (C), manganese (Mn), silicon (Si), chromium (Cr), nickel (Ni), molybdenum (Mo), and copper (Cu), providing a comprehensive dataset for the modeling process.

Steel is heated and tested in accordance with relevant ASTM standards under standard conditions for data collection. The grain size of the steel is measured on the ASTM scale and has been determined to be 7 [78].

Table 3.2 and Table 3.3 show example data collected for five distinct steels: 42CrMo4, 17CrNi6-6, C45E, 20CrMo5 and 16MnCr5.

Table 3.2. Obtained data of the weight percentages of seven alloying elements for five distinct steels

Steel	**Chemical composition, wt%**						
	C	**Mn**	**Si**	**Cr**	**Ni**	**Mo**	**Cu**
42CrMo4	0.41	0.63	0.23	1.01	0.28	0.2	0.14
17CrNi6-6	0.17	0.50	0.28	1.54	1.54	0.08	0.28
C45E	0.48	0.74	0.22	0.28	0.10	0.05	0.18
20CrMo5	0.20	0.93	0.23	1.16	0.09	0.23	0.16
16MnCr5	0.18	1.25	0.31	0.91	0.16	0.04	0.13

Table 3.3. Hardness values in HRC obtained for five distinct steels

Steel	Hardness (HRC) at different distance (in mm) from quenched end												
	1.5	3	5	7	9	11	13	15	20	25	30	40	50
42CrMo4	58	57	56.5	56	55	53	52	50	43	40	37	35	34
17CrNi6-6	43.5	43.5	43	42	41	39	37.5	36	33	32	31.5	31	30.5
C45E	59	57	49	34	31.5	30.5	29.5	29	27	25	23	21	19
20CrMo5	45.5	45	44.5	43	40.5	37.5	35	33.5	31	29.5	28	26.5	24
16MnCr5	44	42.5	39	35	32	30	29	27.5	25.5	24	23	21	19

Table 3.4 outlines the ranges of key predictor parameters for all steels used in modeling steel hardenability.

Table 3.4. Chemical composition ranges for steels

Range (wt.%)	C	Mn	Si	Cr	Ni	Mo	Cu
Minimum	0.12	0.38	0.12	0.09	0.04	0.01	0.07
Maximum	0.7	1.4	0.41	1.97	2.739	0.43	0.34

In addition to chemical composition, the microstructure plays a key role in determining hardness. Therefore, microstructure was also considered in the modeling process. While heat treatment, particularly quenching, is relatively simple to perform, it remains a complex physical process. Even small deviations from the standard procedure can lead to significantly different results. Moreover, regardless of the steel grade or standard, absolute homogeneity in the steel cannot be expected. However, it is nearly impossible to precisely determine the microstructure distribution for specified intervals (e.g., 1.5, 3, 5, 7 mm, etc.) relative to the quenched end of the Jominy specimen. Martensite, being the hardest phase, has the greatest influence on hardness, particularly when present in substantial amounts. A substantial proportion is

generally considered to be 50% or more [78]. Consequently, to minimize errors in the input data, only the presence of martensite was considered. Specifically, for the specified measurement points relative to the quenched end of the Jominy specimen, a martensite content greater than 50% is labeled as 1, while the rest is labeled as 0. Limit of 50% of martensite is determined by the relation. For calculating the hardness value with 50% martensite ($HRC_{50\%M}$), according to maximum hardness, HRC_{max} it is used following relation [79, 80]:

$$HRC\ (M_{50}) = k \cdot HRC_{max} \tag{3.1}.$$

Ratio of hardness of steel with 50% martensite in the microstructure and 99% of martensite in the microstructure (coefficient k) depends on carbon percentage and values for different carbon content is shown in Figure 3.1 [75, 79].

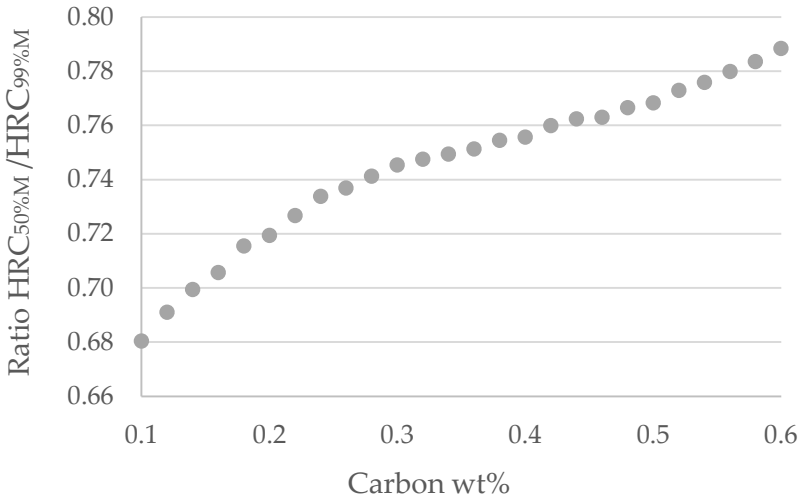


Figure 3.1. Ratio of hardness of steel with 50% martensite in the microstructure and 99% of martensite in the microstructure for different carbon content

For distances on the Jominy specimen where the hardness values exceed $k \cdot \mathrm{HRC}_{max}$, a value of 1 is assigned (indicating that more than 50% of martensite is present in the microstructure). For hardness values lower than $k \cdot \mathrm{HRC}_{max}$, a value of 0 is assigned (indicating that less than 50% of martensite is present in the microstructure).

Example of predictors (indicators of presence of martensite) for steel 42CrMo4 is shown in Figure 3.2 and in Table 3.5.

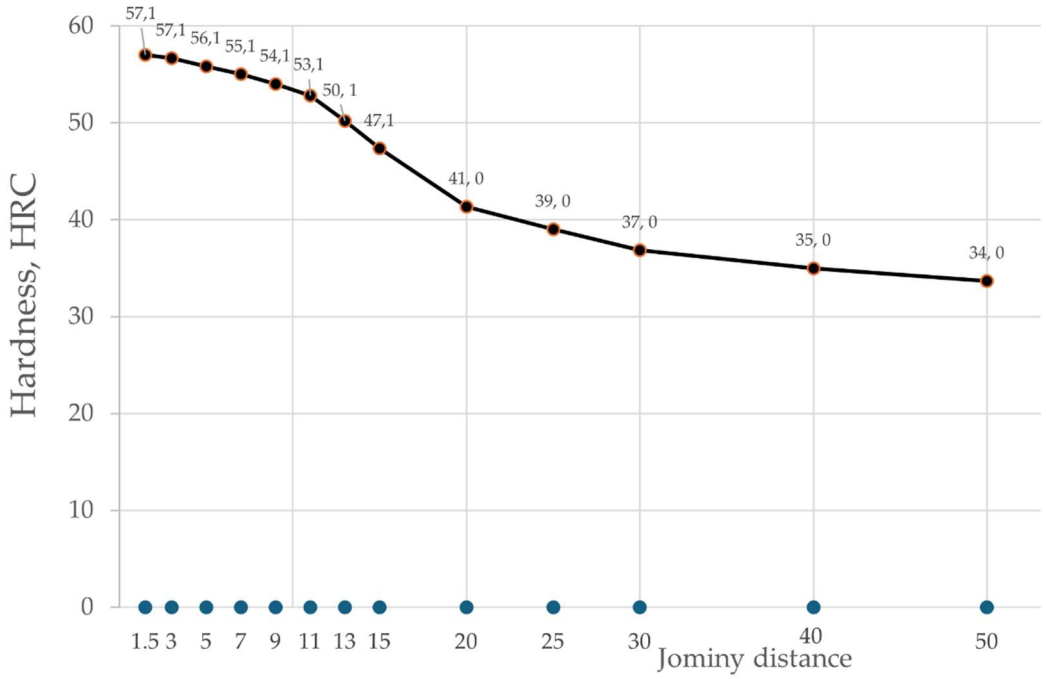


Figure 3.2. Predictors of steel 42CrMo4 (hardnesses and microstructures)

Table 3.5. Predictors for one observation (steel 42CrMo4)

$M_{1.5}$	M_3	M_5	M_7	M_9	M_{11}	M_{13}	M_{15}	M_{20}	M_{25}	M_{30}	M_{40}	M_{50}
1	1	1	1	1	1	1	1	0	0	0	0	0

3.2. Splitting data set

Splitting a dataset into subsets for training and validation is fundamental in machine learning. It plays a key role in tasks such as model evaluation, comparison, and hyperparameter tuning. Common methods include holdout, bootstrap, and cross-validation (CV).

In these methods, the dataset is divided into two subsets: a training set, used to determine the model parameters, and a validation set (test set), which is used to evaluate the model's performance. The training set is crucial for parameterization,

while the validation set serves as an independent reference point for performance. The final model is selected based on its performance on the validation set, ensuring it generalizes well to unseen data.

Cross-validation (CV) is a widely used resampling method in statistical learning. It estimates the test error of a given model and is essential for model evaluation and selection. Cross-validation also helps determine the optimal degree of model flexibility, which is critical for refining and optimizing statistical models.

In *k*-fold cross-validation, the dataset is divided into *k* folds, with each fold used as a validation set once while the others form the training set. Empirical evidence suggests that $k = 10$ often provides a good balance, yielding test error estimates with moderate bias and variance. This helps mitigate the risks of underfitting (high bias) and overfitting (high variance) during model evaluation.

When *k* is too small (e.g., $k = 2$), the training set becomes smaller, which can lead to higher bias. This means the model may underfit the data, as it doesn't have enough training data to capture complex patterns. When *k* is too large, the training set is almost as large as the original dataset, leading to low bias but high variance. This is because the model's performance can vary significantly depending on small changes in the validation set, especially if there are outliers or noise.

Underfitting means the model is too simple to capture the underlying patterns in the data. It doesn’t learn enough from the training set, leading to poor performance on both the training data and new data. This typically happens when the model has too few parameters or is not trained long enough to capture the complexity of the data. The key sign of underfitting is high error on both the training and test data.

Overfitting means the model has learned the training data too well. It performs very well on the training set but struggles to generalize to new, unseen data. The model essentially "memorizes" the training data, capturing noise and outliers rather than

generalizable patterns. The key sign of overfitting is low training error (loss) but high-test error (loss) (Figure 3.3).

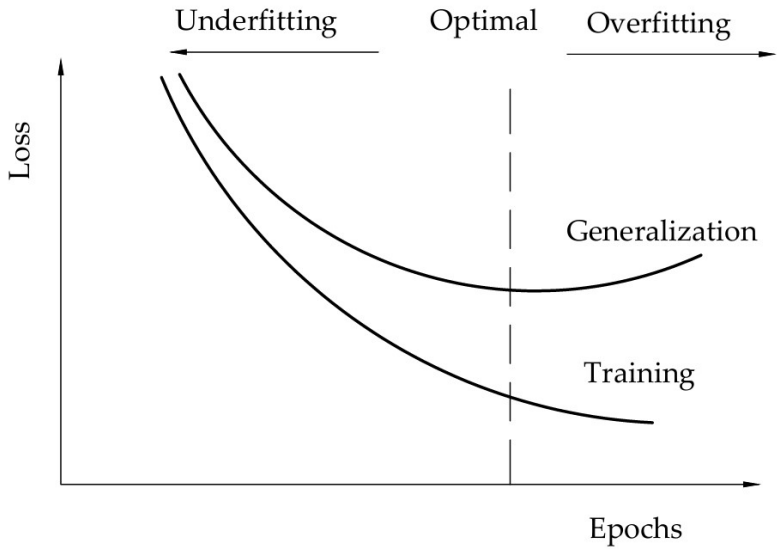


Figure 3.3. Illustration of neural network learning with early stopping technique

Four approaches to organizing the dataset were applied in an evolving sequence, each based on the analysis of the dataset, which included features such as Jominy test results (hardness), microstructure (indicating the presence of more than 50% martensite), and the mass fraction of seven alloying elements for various steels. The details of these approaches were presented in Sections 3.3 to 3.6. These approaches were chosen to address specific challenges in modeling and data representation, with each method progressively refining the dataset for improved analysis. The purpose of this approach was to maximize the use of these features, eventually producing the most efficient model for designing the chemical composition of steels based on the required hardenability. The results for each ANN model, based on the different dataset organizations, were presented in Section 4.

3.3. Organization of the Dataset Approach 1

The input data for each steel are structured as shown in Table 3.6. The dataset consists of 14 columns—13 for hardness measurements and 1 for microstructure—and 13 rows, each corresponding to a specific Jominy distance.

Table 3.6. First approach at data structuring – Example of input data (predictors) organized for one steel (C45E)

Predictors													
59	0	0	0	0	0	0	0	0	0	0	0	0	1
0	57	0	0	0	0	0	0	0	0	0	0	0	1
0	0	49	0	0	0	0	0	0	0	0	0	0	1
0	0	0	34	0	0	0	0	0	0	0	0	0	0
0	0	0	0	31.5	0	0	0	0	0	0	0	0	0
0	0	0	0	0	30.5	0	0	0	0	0	0	0	0
0	0	0	0	0	0	29.5	0	0	0	0	0	0	0
0	0	0	0	0	0	0	29	0	0	0	0	0	0
0	0	0	0	0	0	0	0	27	0	0	0	0	0
0	0	0	0	0	0	0	0	0	25	0	0	0	0
0	0	0	0	0	0	0	0	0	0	23	0	0	0
0	0	0	0	0	0	0	0	0	0	0	21	0	0
0	0	0	0	0	0	0	0	0	0	0	0	19	0

In this data structure, the original 470 observations (samples) are transformed into 6110 observations (470 × 13). Each steel sample is represented by 13 rows in total. A microstructure indicator (values 1 or 0) is added as the 14th column, making the input data consist of 14 features in total (13 hardness values and 1 corresponding to microstructure). The predictors (input data) comprise 85540 values (14 features × 6110 observations).

The responses (output data) include 42770 values (7 features × 6110 observations). The output data is structured as shown in Table 3.7.

Table 3.7. Example of output data (responses) for steel C45E

C	Mn	Si	Cr	Ni	Mo	Cu
0.48	0.74	0.22	0.28	0.10	0.05	0.18

3.4. Organization of the Dataset Approach 2

The next approach focused on analyzing each specified distance relative to the quenched end of the Jominy specimen. For each steel, data is organized across 13 specified measurement points of Jominy specimen - distances from the quenched end of the Jominy specimen. For each measurement points of Jominy specimen, three parameters were included:

- Jominy distance (J_d),
- Hardness (HRC),
- Microstructure (M).

A microstructure indicator (values 1 or 0) remained the same as in Approach 1. For each steel sample, data is collected across 13 specified distances from the quenched end, characterized by 39 parameters (J_d, HRC, and M). The output for each steel is the mass fraction of 7 selected alloying elements.

The data arranged in this structure are illustrated in Table 3.8.

Table 3.8. Second approach at data structuring - Input data (predictors) for first 6 points of Jominy curve for steel C45E

Predictors																	
1.5	59	1	3	57	1	5	49	1	7	34	0	9	31.5	0	11	30.5	0

In this structure, the total number of observations remains consistent with the original version, with 470 observations in total. This approach simplifies the data organization compared to the first attempt.

Predictors (input data) consist of 39 features (J_d, HRC, M) × 470 observations (for each steel). Responses (output data) consist of seven features (the mass fraction of the seven chemical elements) × 470 observations (for each steel) – the same as in the attempt no.1.

By organizing the data this way, the dataset maintains its original number of observations but in a more manageable and structured format, which facilitates easier analysis and modeling.

3.5. Organization of the Dataset Approach 3

To simplify the data structure and reduce the number of input features, a new approach is considered. In this method, one parameter is 'fixed,' which is the Jominy distance (J_d). Although there are 13 different Jominy distances, each hardness value and microstructure are specifically associated with a particular Jominy distance (J_d).

For example, the first hardness value is always associated with J_d = 1.5 mm, and the fifth hardness value is associated with J_d = 9 mm. This results in a reduction of input features: instead of 39 features, only 26 features are considered.

The data arranged in this structure are illustrated in Table 3.9.

Table 3.9. Input data (predictors) for all 13 points of Jominy curve for one steel (C45E)

Predictors																									
59	1	57	1	49	1	34	0	31.5	0	30.5	0	29.5	0	29	0	27	0	25	0	23	0	21	0	19	0

Predictors (input data) consist of 26 features × 470 observations. Responses (output data) remain the same, consisting of 7 features × 470 observations.

3.6. Organization of the Dataset Approach 4

The input data are structured in the same way as in the previous instance (Dataset Approach 3) and are presented in Table 3.10. The key distinction lies in the output data (responses). Each alloying element is treated as an individual output, as neural networks typically have a single output. Consequently, seven separate neural network models are developed, one for each alloying element

Table 3.10. Input data (predictors) for one steel (65Mn4)

Predictors																									
64	1	63	1	59.5	1	41	0	36	0	36	0	35.5	0	35	0	35	0	33.5	0	32	0	30	0	27.5	0

Considering that there are 26 input data points and one output data point (chemical element), the neural networks architecture will consist of 26 input nodes and a single output node. Seven neural network models, one for each of the seven alloying elements, need to be developed.

The first step in modeling individual neural networks for each alloying element is to divide the dataset into a training dataset and a test dataset.

Data partitioning is done in MATLAB 2023b using the *cvpartition* function. This function generates a random, non-stratified partition for holdout validation on a dataset. The proportion of observations assigned to the test set is set to 5%, while the remaining data form the training set.

By dividing a dataset, the objective is to develop a model using the training dataset and evaluate its performance on the test dataset. For optimal results, both the training

dataset and test dataset should be representative samples of the overall data distribution. This ensures that the model learns meaningful patterns during training and is evaluated reliably on the test dataset. Otherwise, the model may not generalize well. The Kolmogorov–Smirnov (K-S) test assesses whether the training and test datasets follow the same distribution by detecting significant differences between them [80]. The null hypothesis of the K-S test states that both datasets come from the same underlying distribution.

The Kolmogorov–Smirnov (K-S) test values, representing the maximum absolute differences between the cumulative distribution functions, were calculated for the features hardness (HRC) and microstructure (M) at specific Jominy distances. The test was not performed for Jominy distances because their distribution is the same.

The results of the Kolmogorov-Smirnov test for the hardness (HRC) feature and microstructure at specified distances from the quenched end of the Jominy specimen are presented in Tables 3.11 and 3.12, respectively.

Table 3.11. Kolmogorov-Smirnov test results of the training and test dataset

Features	**$HRC_{1.5}$**	**HRC_{3}**	**HRC_{5}**	**HRC_{7}**	**HRC_{9}**	**HRC_{11}**	**HRC_{13}**	**HRC_{15}**	**HRC_{20}**	**HRC_{25}**	**HRC_{30}**	**HRC_{40}**	**HRC_{50}**
KS -test Value	0.152	0.163	0.166	0.127	0.093	0.090	0.147	0.161	0.157	0.132	0.132	0.167	0.147
p-value	0.68	0.6	0.57	0.86	0.99	0.99	0.72	0.61	0.64	0.82	0.84	0.57	0.73

Table 3.12. Kolmogorov-Smirnov test results of the training and test dataset

Features	**$M_{1.5}$**	**M_{3}**	**M_{5}**	**M_{7}**	**M_{9}**	**M_{11}**	**M_{13}**	**M_{15}**	**M_{20}**	**M_{25}**	**M_{30}**	**M_{40}**	**M_{50}**
KS -test Value	0	0	0.018	0.106	0.071	0	0.001	0.037	0.031	0.027	0.044	0.037	0.009
p-value	1	1	1	0.96	1	1	1	1	1	1	1	1	1

The K-S test values obtained for hardness (HRC) range from 0.09 to 0.161, while those for microstructure (M) range from 0 to 0.106. These values suggest relatively small differences between the distributions. Additionally, *p*-values for each test ranged from 0.66 to 0.99 for hardness and from 0.9 to 1.0 for microstructure, indicating no

substantial evidence to reject the null hypothesis of similarity between the distributions.

These results further support the similarity of feature distributions between the training and test datasets, providing confidence in the model's ability to generalize well to unseen test data (Figure 3.4).

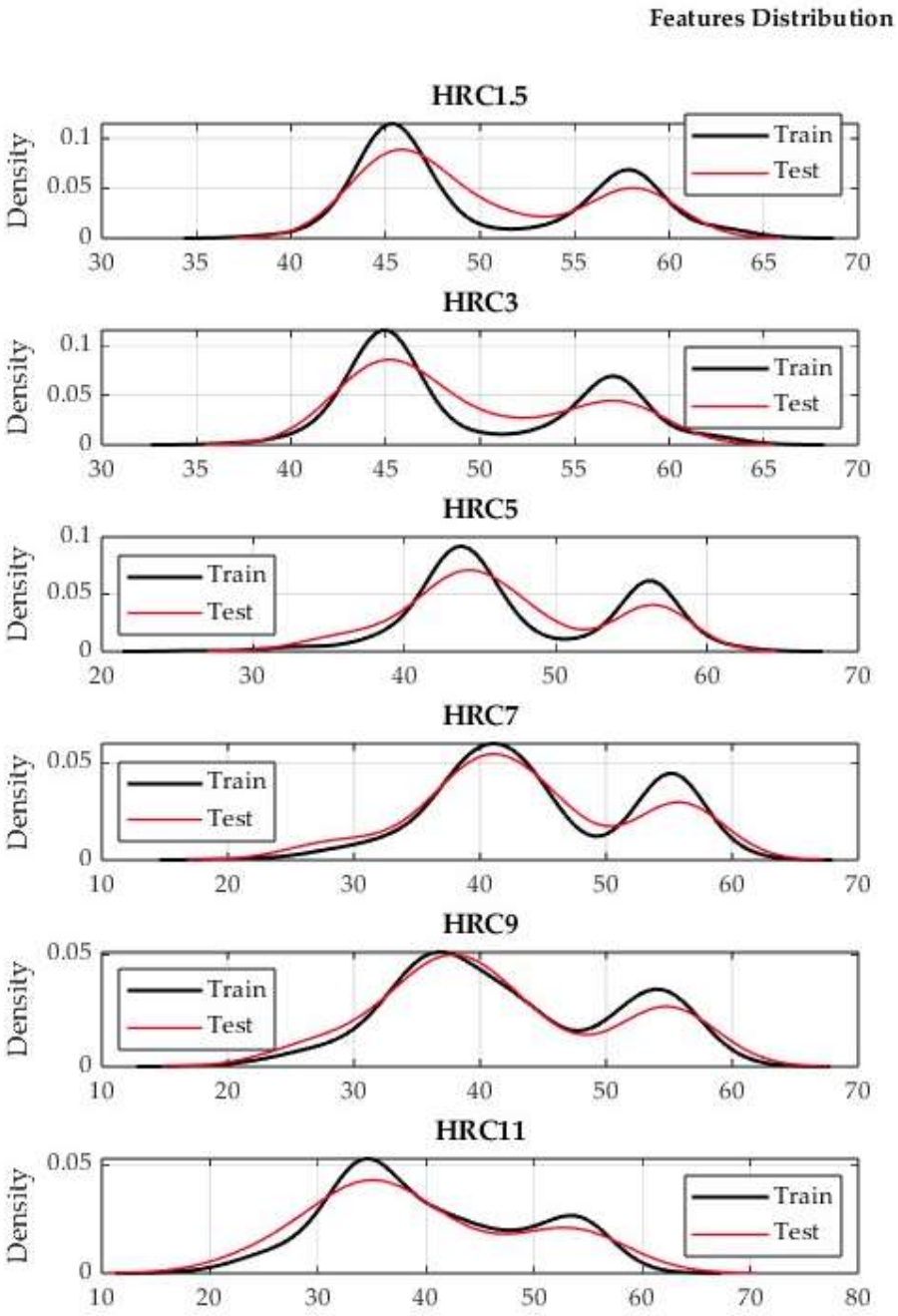


Figure 3.4. Training and test datasets' split distribution for first six nearest distances from quenched end of Jominy specimen

4. RESULTS OF THE ANN APPLICATION FOR VARIOUS DATA REPRESENTATION

For a collected dataset supervised artificial neural networks are used to model complex regression relationships. To design artificial neural networks, application called Neural Net Fitting from MATLAB version 2023b is used. Deferent architectures of artificial neural network were tested.

MATLAB 2023b was utilized for regression development, with the *trainlm* training function employed. *Trainlm* is a network training function that updates weight and bias values using Levenberg-Marquardt optimization. The backpropagation method, known for its efficiency, is recommended as the primary supervised learning algorithm [81].

The Levenberg-Marquardt (LM) optimization is an iterative method used to solve non-linear least squares problems. It combines the advantages of two optimization techniques: gradient descent and the Gauss-Newton method. When the solution is far from optimal, LM behaves like gradient descent, taking cautious, small steps to prevent divergence. As the solution approaches the optimal point, it transitions to the faster Gauss-Newton method for fine-tuning [82]. LM is designed to converge faster than standard gradient descent by using a second-order approximation of the error surface, similar to Newton's method.

Gradient Descent relies on the first derivative of the error function to find the direction of the steepest descent. It adjusts the parameters incrementally in that direction:

$$w_{new} = w_{old} - \eta \cdot \nabla E \qquad (4.1),$$

where η is the learning rate, and ∇E is the gradient of the error with respect to the parameters. Gradient descent assumes that the error surface is linear in the region it's working with, which can lead to slow convergence when the surface is highly curved or when the learning rate is poorly tuned.

Newton's method incorporates the second derivative of the error surface, represented by the Hessian matrix, *H*. Newton's update rule is:

$$w_{new} = w_{old} - H^{-1} \cdot \nabla E \tag{4.2}.$$

The Hessian matrix, *H*, contains second-order partial derivatives of the error function and provides information about the curvature of the error surface. Using H^{-1} allows the algorithm to adjust the step size dynamically, taking larger steps in shallow regions and smaller steps in steep regions.

Hessian matrix is defined as:

$$H = \nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{x_1^2} & \cdots & \frac{\partial^2 f}{x_1 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{x_n x_1} & \cdots & \frac{\partial^2 f}{x_n^2} \end{pmatrix} \tag{4.3}.$$

The Gauss-Newton method is based on the Newton-Raphson method, but specially adapted for solving the nonlinear least squares problem. Namely, unlike Newton's method, Gauss-Newton's method does not use other derivatives, but approximates the Hesse matrix that is needed for calculation.

According to the Gauss-Newton's method the approximation of expression 4.3 is:

$$H \approx 2J^T J \tag{4.4},$$

Where J is Jacobian matrix.

Jacobian matrix is:

$$J = \begin{bmatrix} \frac{\partial f_1}{x_1} & \cdots & \frac{\partial f_1}{x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{x_1} & \cdots & \frac{\partial f_n}{x_n} \end{bmatrix} \tag{4.5},$$

The gradient is computed as:

$$\nabla E = J^T e \quad (4.6),$$

where e is is the vector of errors.

LM updates the weights using the following formula:

$$w_{i+1} = w_i - (J^T J + \mu I)^T J^T e \quad (4.6),$$

where $J^T J$ approximates the Hessian matrix, μI adds a damping factor to ensure numerical stability, where μ is a positive scalar, and I is the identity matrix.

When the solution is far from the minimum, LM increases μ, making $(J^T J + \mu I)$ dominate, causing LM to behave like gradient descent. As the solution nears the minimum, μ is reduced, and LM transitions to the Gauss-Newton method. Near the minimum steps are efficient and decrease in size due to diminishing residuals and gradients.

The model with the smallest test root mean square error (RMSE) is considered the best model, as a low RMSE indicates that the selected model is likely to generalize well to new datasets. In addition to RMSE, the model's performance is evaluated using other metrics, including mean square error (MSE), **m**ean absolute error (MAE), and the coefficient of determination (R^2), during the test performance analysis [83].

The Root Mean Square Error (RMSE) is calculated as:

$$RMSE_j = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4.7),$$

where j states as the j-th iteration (of 10), n is total number of observations, y_i is the response (actual) value of the i-th observation and $\hat{y}_i$ is the predicted value of the i-th observation.

During the application of 10-fold cross-validation, the Root Mean Squared Error (RMSE) is computed for each of the 10 iterations, and the mean of these values is reported as the model's RMSE:

$$RMSE = \frac{1}{10}\sum_{j=1}^{10} RMSE_j \quad (4.8).$$

The coefficient of determination, denoted as R^2, is defined as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (4.9),$$

where SS_{res} is residual sum of squares and it is defined as: $SS_{res} = \sum_i (y_i - \hat{y}_i)^2$, where y_i are actual values, while $\hat{y}_i$ are predicted values, and SS_{tot} is total sum of squares and it is defined as $SS_{res} = \sum_i (y_i - \bar{y})^2$, where y_i are actual values, while $\overline{y_i}$ is mean of actual values [84, 85].

The coefficient of determination is usually between values 0 and 1. When the coefficient of determination is equal to 0, model explains no variability and predictions are no better than the mean ($SS_{res} = SS_{tot}$). The perfect fit is when the coefficient of determination is equal to 1. The model's predictions perfectly match the actual data points, leaving no unexplained variation ($SS_{res} = 0$)[3].

A negative R^2 can occur in models that fit worse than a simple horizontal line at $\bar{y}$ ($SS_{res} > SS_{tot}$) [86].

[3] Unexplained variability is the part of the variability that the model cannot capture. This is the Residual Sum of Squares (SS_{res}).

4.1. Results of the ANN application for Dataset Approach 1

The neural network architecture used in this study is denoted as [14-27-27-7], indicating an input layer with 14 nodes, two hidden layers each containing 27 nodes, and an output layer with 7 nodes corresponding to 7 chemical elements.

As described in Section 3.2, the dataset is randomly divided into three subsets: the training dataset (70% of the total data), the validation dataset (15% of the total data), and the test dataset (15% of the total data). Prior to training, the input data (predictors) and output data (responses) were normalized to a specific range using the *mapminmax* procedure in MATLAB [87]. The training function utilized is *trainlm*, which employs Levenberg-Marquardt optimization. Plot of training results are shown in Figure 4.1.

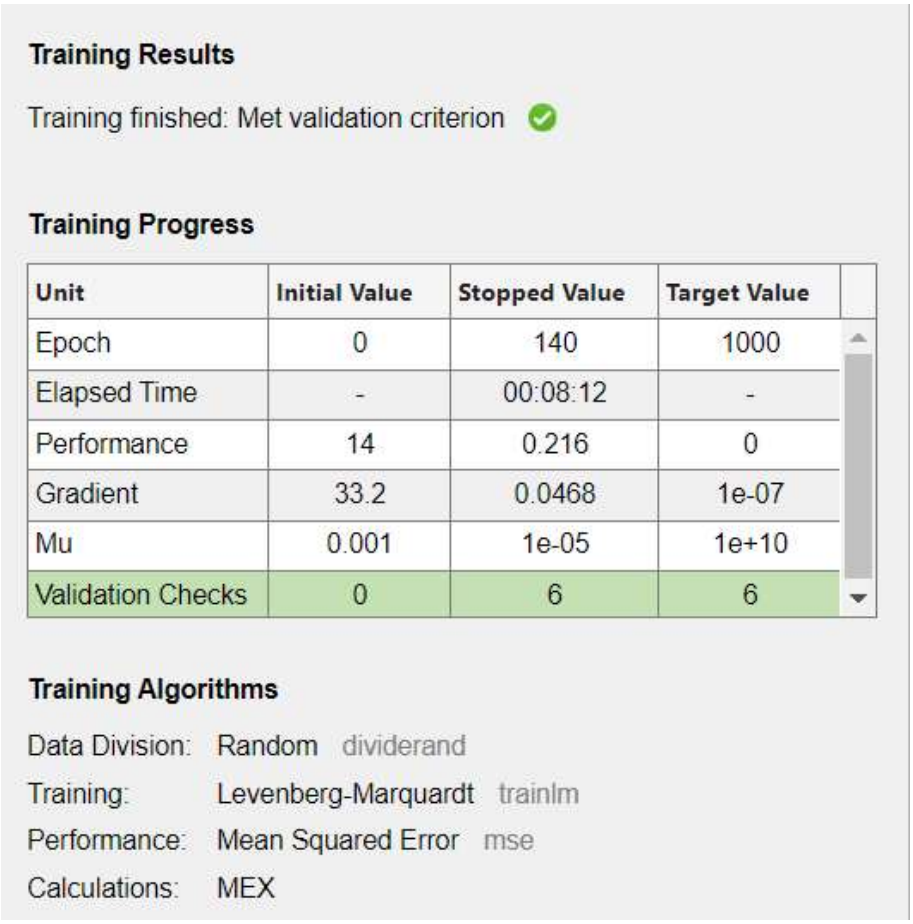


Figure 4.1. Training results Dataset Approach 1

The maximum number of training epochs is set to the default value of 1000. The initial value of μ (the damping parameter) is set to its default value of 0.001, while the target value of μ is set to the default value of 1000. In the Levenberg-Marquardt (*trainlm*) algorithm, the damping term is represented as μI, where I is the identity matrix. A very large μ makes *trainlm* behave like slow gradient descent and can cause stagnation. At the start of training, the gradient is computed based on the initial randomly assigned weights and is found to be 33.2. The target gradient value is set to the default value of 10^{-7}. If the gradient is smaller than this value, training is considered converged, and the algorithm should stop. Maximum validation failures. is set to the default value of 6. The validation error is a measure of how well the model performs on a separate validation dataset. The condition specifies that if no improvement in the validation error occurs for six consecutive validation checks (epochs), the model has not made significant progress in terms of generalization, and training should stop.

The training process is structured so that it continues until one of the stopping criteria is met. In this case, training proceeded until the validation error was greater than or equal to the previously recorded minimum validation error for six consecutive validation iterations ("Met validation criterion").

Neural network training performance is shown in Figure 4.2

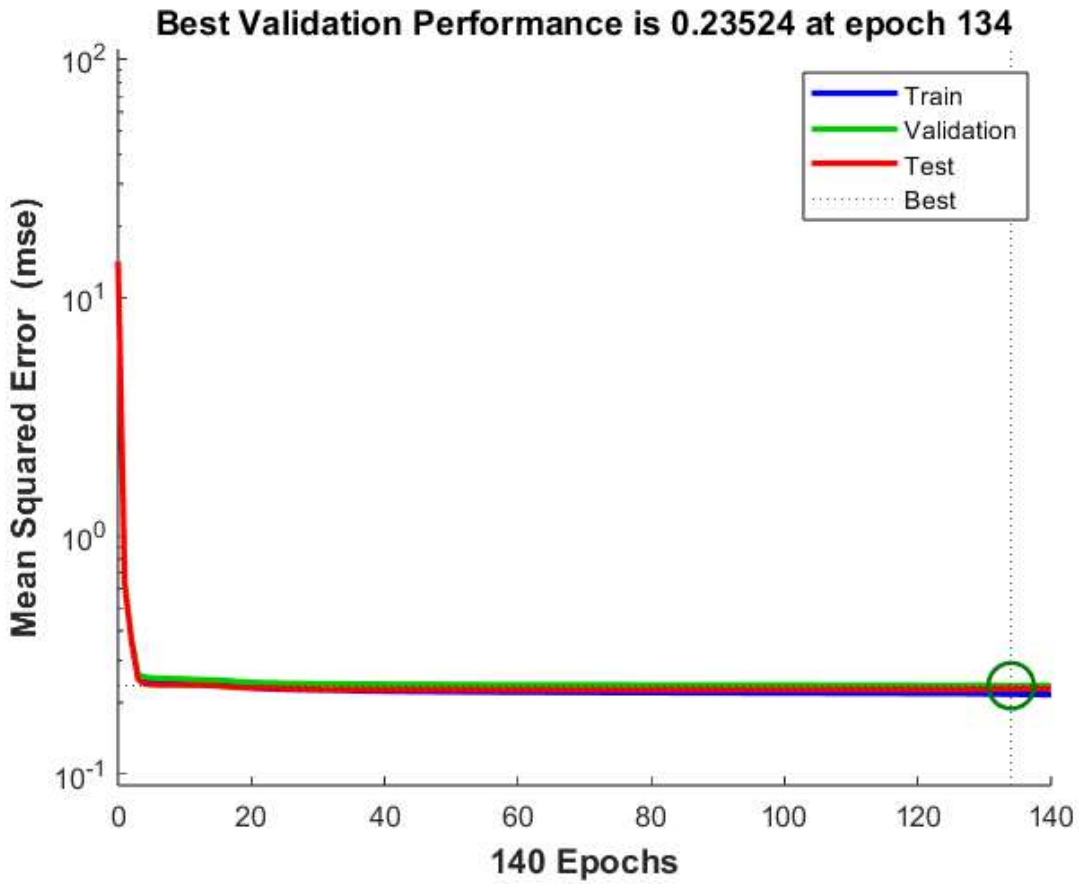


Figure 4.2. Neural network training performance for Dataset Approach 1

The best validation performance of MSE=0.235 was achieved at epoch 134. The training continued for six more epochs before stopping. The best validation performance corresponds to RMSE=0.485.

Generally, the error decreases after more epochs of training, but it may start to increase on the validation dataset as the network begins to overfit the training data, as shown in Figure 3.3. In Figure 4.2, this was not the case, as the gap between the training and test datasets is too small, which is undesirable. A small gap typically indicates underfitting, suggesting that the model is not learning effectively from the training data.

The regression plot in Figure 4.3 displays the network predictions (output) with respect to responses (target) for the training, validation, and test sets.

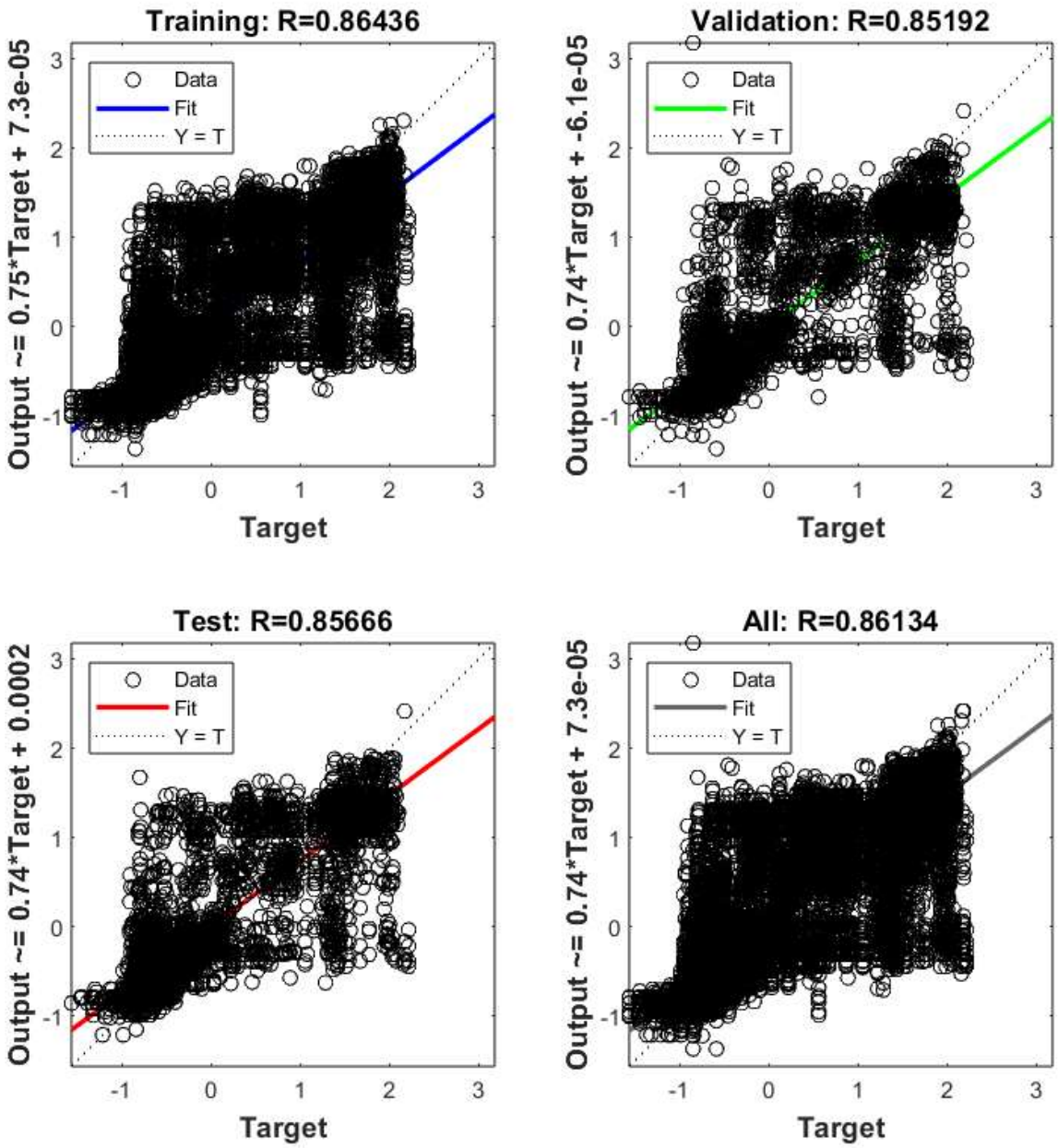


Figure 4.3. Comparison between the expected and actual response plots for Dataset Approach 1

The correlation coefficient, R for the test dataset is 0.86, corresponding to a coefficient of determination, R^2 of 0.74. However, the ANN model produces physically

meaningless results. Therefore, further optimization, including improved data splitting and architectural adjustments, has been discontinued.

4.2. Results of the ANN application for Dataset Approach 2

The input layer consists of 39 nodes, incorporating 13 distances from the quenched end of the Jominy specimen, 13 corresponding hardness values, and 13 indicators of martensite presence in the microstructure at those distances. The output layer comprises seven nodes, each corresponding to a specific chemical element. The proposed neural network architecture is [39-14-14-7], featuring two hidden layers, each with 14 nodes. The dataset is randomly divided into three subsets as previously described. The training function utilized is *trainlm*, which employs Levenberg-Marquardt optimization. Plot of training results are shown in Figure 4.4.

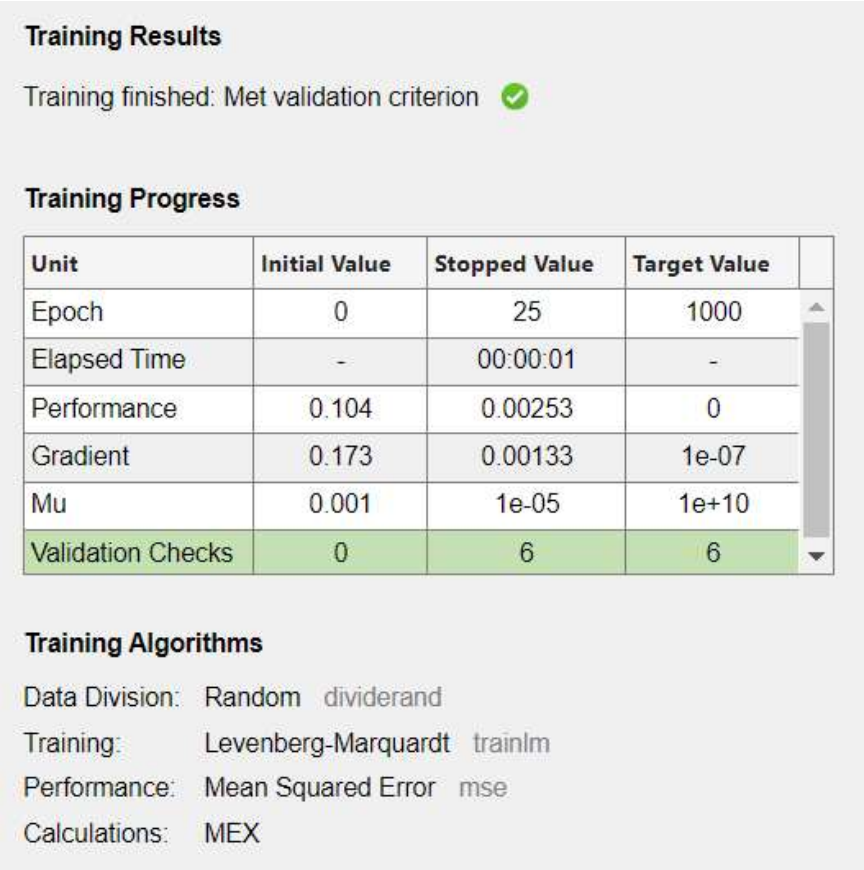


Figure 4.4. Training results Dataset Approach 2

The maximum number of training epochs is set to the default value of 1000. The initial value of μ (the damping parameter) is set to its default value of 0.001, while the target value of μ is set to the default value of 1000. At the start of training, the gradient is computed based on the initial randomly assigned weights and is found to be 0.173. The target gradient value is set to the default value of 10^{-7}. Maximum validation failures. is set to the default value of 6.

The training process is structured so that it continues until one of the stopping criteria is met. In this case, training proceeded until the validation error was greater than or equal to the previously recorded minimum validation error for six consecutive validation iterations ("Met validation criterion").

Neural network training performance is shown in Figure 4.5.

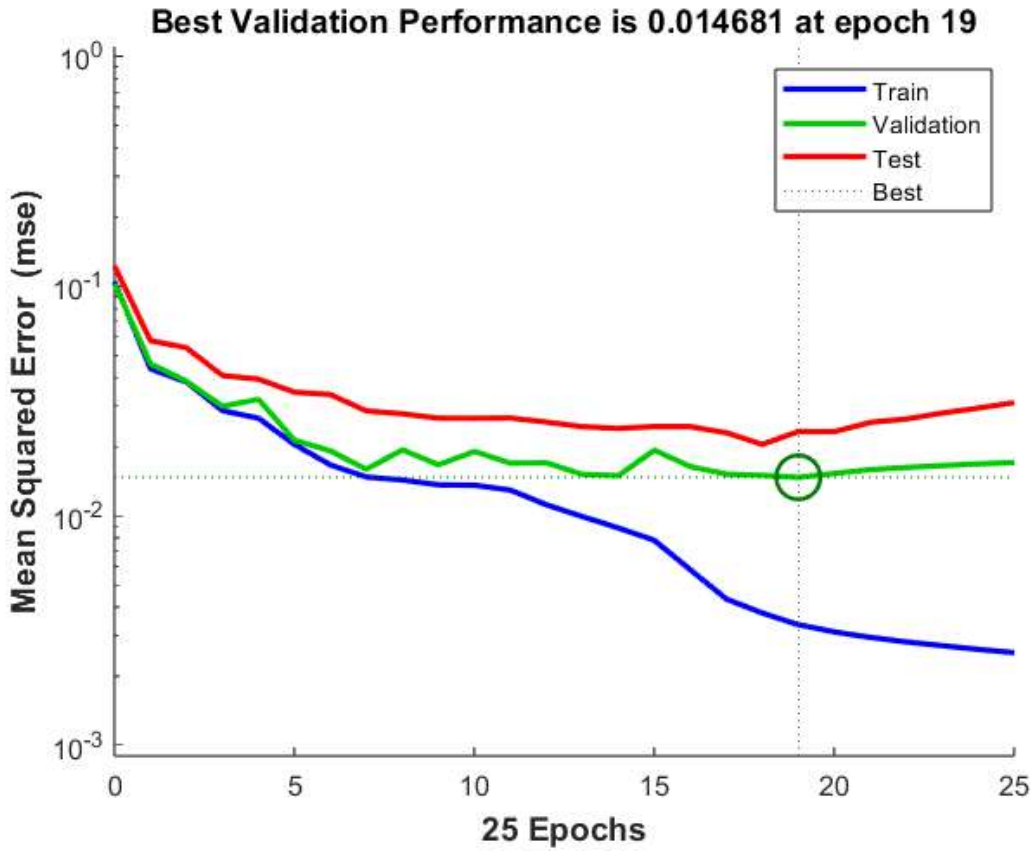


Figure 4.5. Neural network training performance for Dataset Approach 2

The best validation performance of MSE=0.015 was achieved at epoch 19. The training continued for six more epochs before stopping. The best validation performance corresponds to RMSE=0.121.

The regression plot in Figure 4.6 displays the network predictions (output) with respect to responses (target) for the training, validation, and test sets.

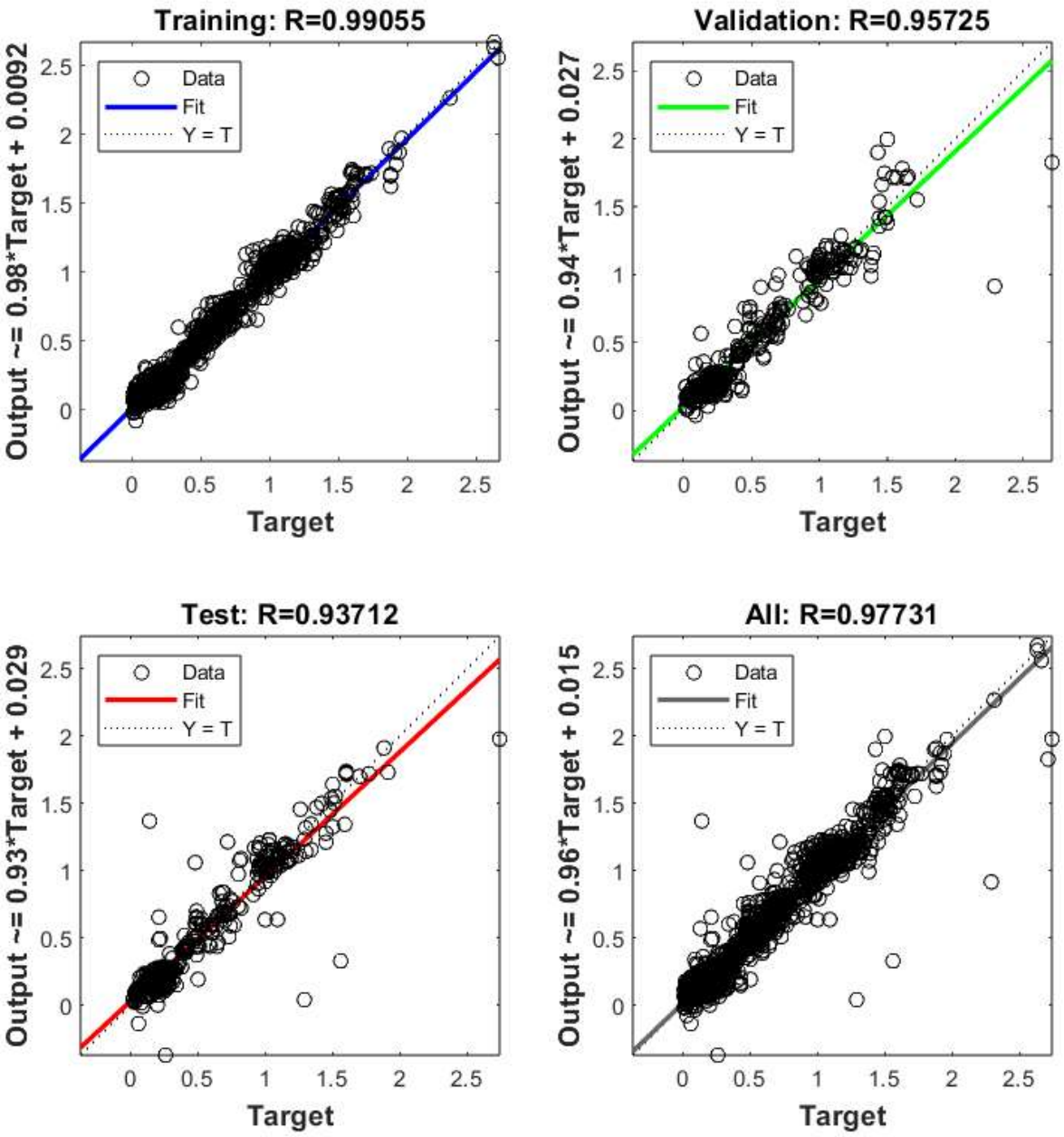


Figure 4.6. Network predictions (output) with respect to responses (target) for the training, validation, and test sets for Dataset Approach 2

The correlation coefficient for the test dataset is R=0.94 which corresponds to coefficient of determination R^2=0.88.

4.3. Results of the ANN application for Dataset Approach 3

The proposed neural network architecture is [26-14-14-7], where the input layer consists of 26 nodes, representing 13 hardness values at various distances from the quenched end of the Jominy specimen and corresponding indicators of martensite presence in the microstructure. The dataset is randomly divided into three subsets, as previously described. The training function utilized is *trainlm*, which employs Levenberg-Marquardt optimization. Plot of training results are shown in Figure 4.7.

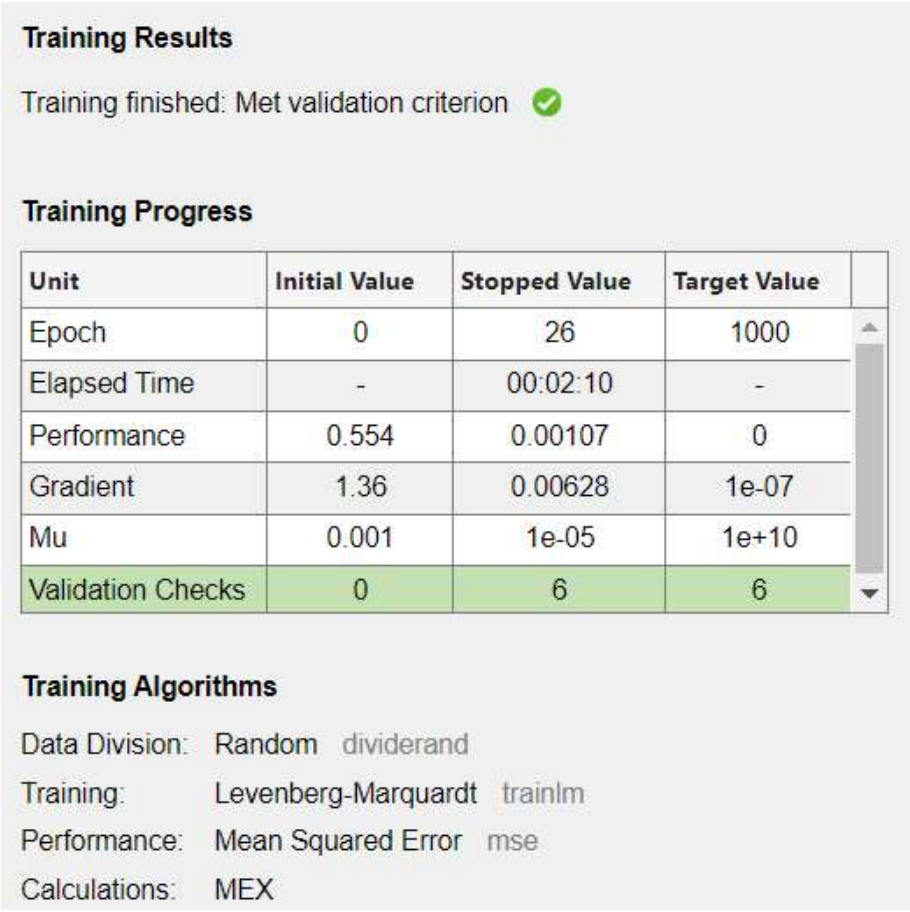


Figure 4.7. Training results Dataset Approach 3

The maximum number of training epochs is set to the default value of 1000. The initial value of μ (the damping parameter) is set to its default value of 0.001, while the target value of μ is set to the default value of 1000. At the start of training, the gradient is computed based on the initial randomly assigned weights and is found to be 1.36. The target gradient value is set to the default value of 10^{-7}. Maximum validation failures. is set to the default value of 6.

The training process is structured so that it continues until one of the stopping criteria is met. In this case, training proceeded until the validation error was greater than or equal to the previously recorded minimum validation error for six consecutive validation iterations ("Met validation criterion"). Neural network training performance is shown in Figure 4.8.

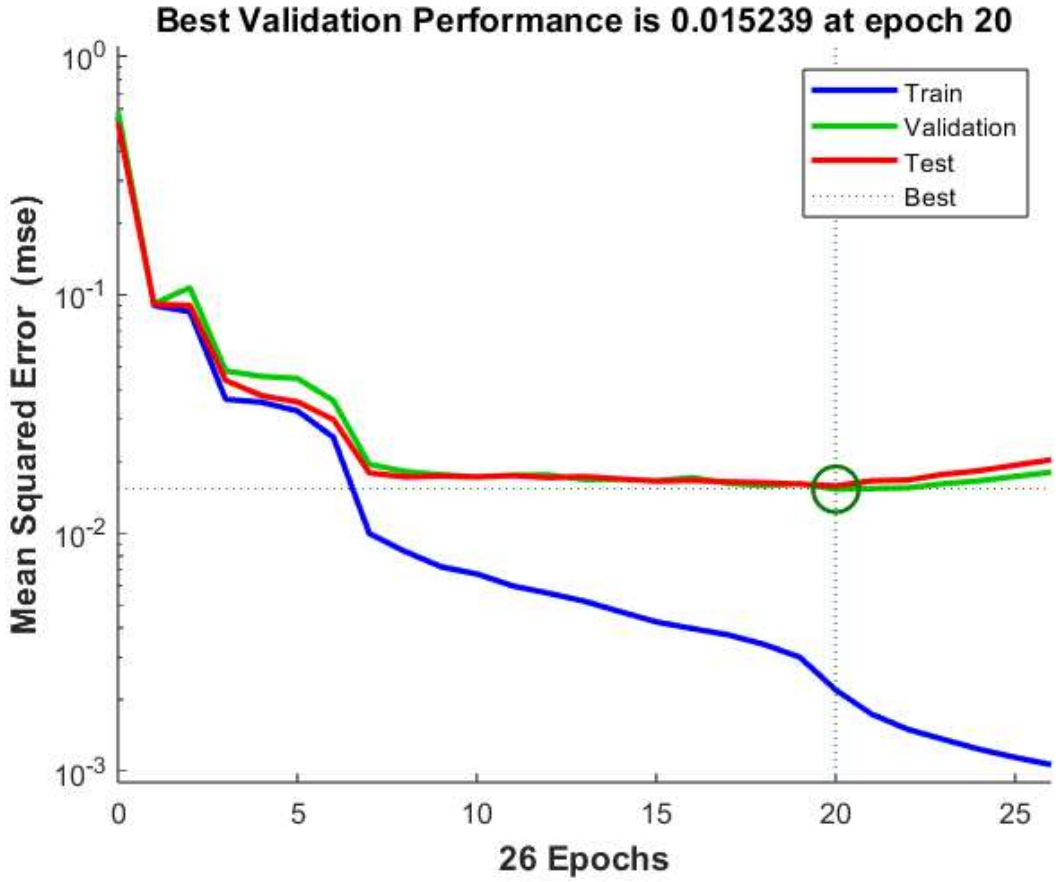


Figure 4.8. Neural network training performance for Dataset Approach 3

The best validation performance of MSE=0.015 was achieved at epoch 20. The training continued for six more epochs before stopping. The best validation performance corresponds to RMSE=0.123.

The regression plot in Figure 4.9 displays the network predictions (output) with respect to responses (target) for the training, validation, and test sets.

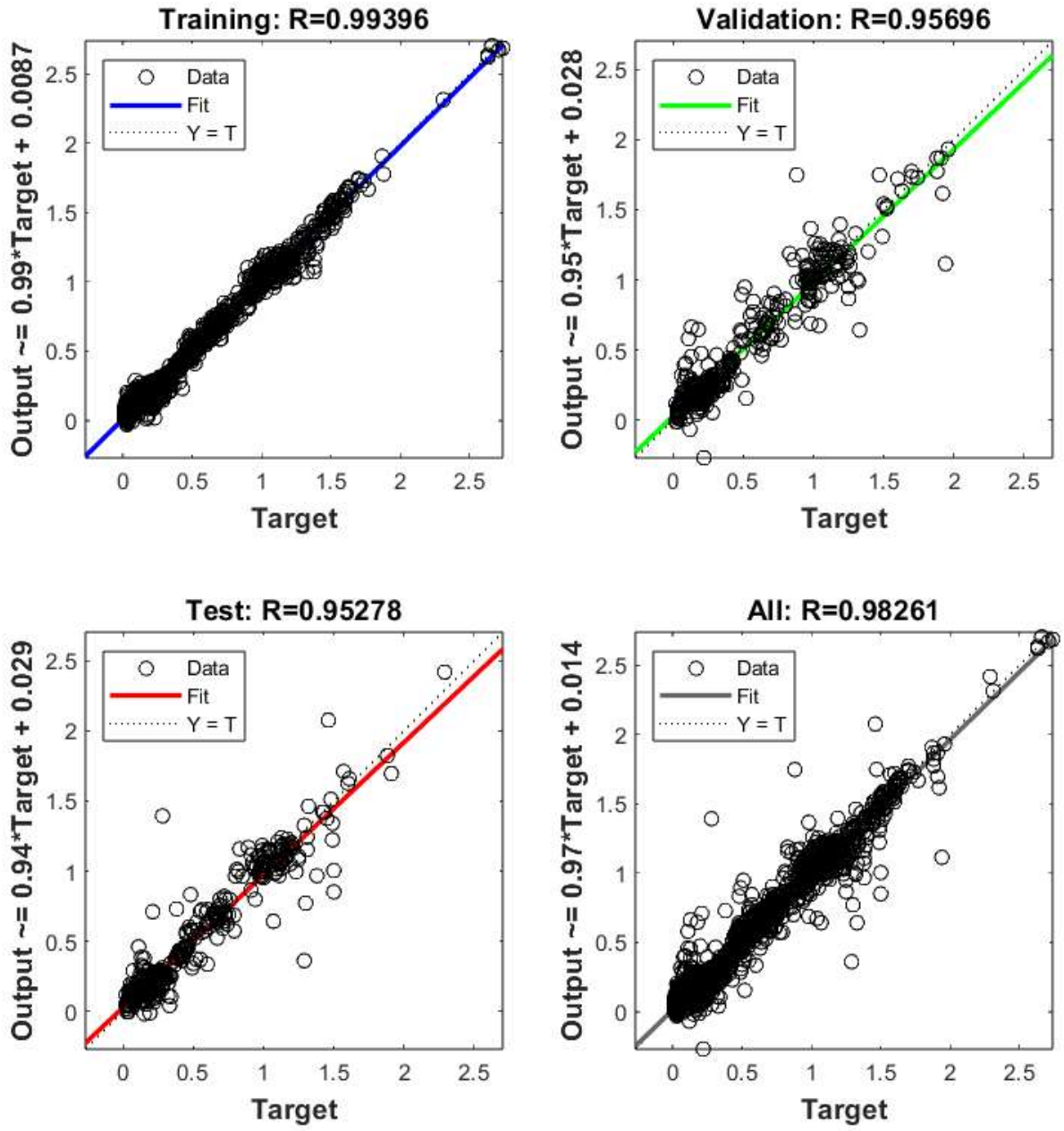


Figure 4.9. Network predictions (output) with respect to responses (target) for the training, validation, and test sets for Dataset Approach 3

The correlation coefficient (R) indicates results that are comparatively favorable, marginally exceeding those obtained with Dataset Approach 2. The correlation coefficient for the test dataset is R=0.95 which corresponds to coefficient of determination R^2=0.90.

4.4. Results of the ANN application for Dataset Approach 4

The Machine Learning and Deep Learning Toolbox from MATLAB R2023b was used for regression tasks. Neural network regression models are trained using the Regression Learner App, with 10-fold cross-validation applied during training to prevent overfitting [88, 89].

The model development procedure, including both the training and testing phases, is carried out independently for each of the seven alloying elements. Five distinct artificial neural network architectures were used for training: 'trilayered', 'bilayered', 'narrow', 'medium', and 'wide' networks, along with an optimized neural network. The specific structures of these neural networks are as follows:

- trilayered [26-10-10-10-1],
- bilayered [26-10-10-1],
- narrow [26-10-1],
- medium [26-25-1],
- wide [26-100-1].

All neural networks had an input layer with 26 nodes and an output layer with one node. The hidden layers vary across the different architectures. Specifically:

The narrow network consists of one hidden layer with 10 nodes. The medium network has one hidden layer with 25 nodes. The broad network includes one hidden layer with 100 nodes. The bilayered network has two hidden layers, each containing 10 nodes. The trilayered network features three hidden layers, each with 10 nodes.

The rectified linear unit (ReLU) activation function is used for all five neural network architectures. For the optimized neural network, both the architecture and activation function are adjusted for improved performance.

The following subsections provide comprehensive reviews of model evaluations and selections, performance analyses, and training procedures to identify the optimal models for predicting the chemical compositions of steels. The training was conducted individually for each chemical element. The alloying elements were analyzed individually, as the distinct impact of each element is essential and often adequate to achieve the desired chemical composition of steel. Consequently, each neural network comprises an input layer with 26 nodes (13 hardness measurements at distances of 1.5, 3, 5, 7, 9, 11, 13, 15, 20, 25, 30, 40 and 50 mm from the quenched end of the Jominy specimen, together with 13 indicators of martensite presence) and an output layer featuring a single node (alloying element).

4.4.1. ANN model for carbon

The artificial neural network models are listed in Table 4.1 and sorted according to their lowest RMSE values for the test results

Table 4.1. Results after training for carbon with six different architectures of ANNs

Model	Architecture	RMSE		R^2		MAE	MSE
		Train	Test	Train	Test	Test	Test
Narrow NN	26-10-1	0.0204	0.0108	0.973	0.994	0.0093	0.0001
Bilayered NN	26-10-10-1	0.0191	0.0114	0.976	0.994	0.0085	0.0001
Trilayered NN	26-10-10-10-1	0.0218	0.0140	0.969	0.991	0.0111	0.0002
Medium NN	26-25-1	0.0200	0.0199	0.974	0.981	0.0137	0.0004
Wide NN	26-100-1	0.0292	0.0290	0.945	0.959	0.0164	0.0008
Optimizable NN	26-4-4-14-1	0.0232	0.0307	0.965	0.954	0.0202	0.0010

The Narrow Neural Network model yielded the best results, featuring one hidden layer with 10 nodes. For this model, the comparison between the predicted responses and the experimental data for each observation from the training dataset is shown in Figure 4.10. Similarly, the predicted responses versus the experimental unseen data (the test dataset) are presented in Figure 4.11. Additionally, Figure 4.11 includes the relationship between the residuals (the differences between predicted responses and experimental data) and the real data. The best neural network model demonstrates excellent performance, with very low RMSE and an exceptionally high R^2 value.

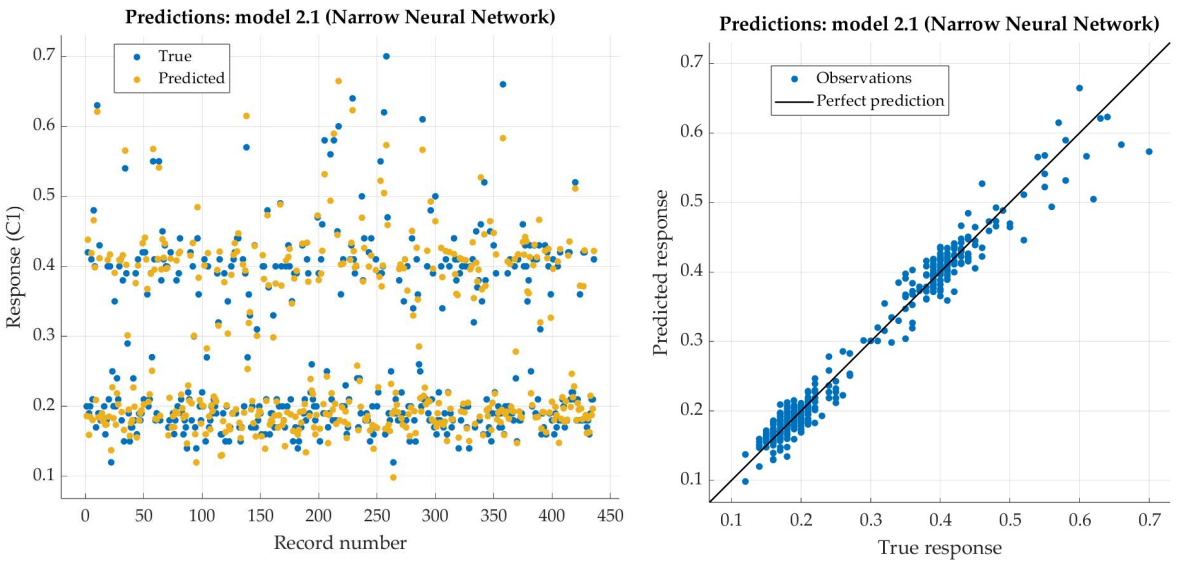


Figure 4.10. Performance evaluation of Narrow Neural Network on training dataset (carbon)

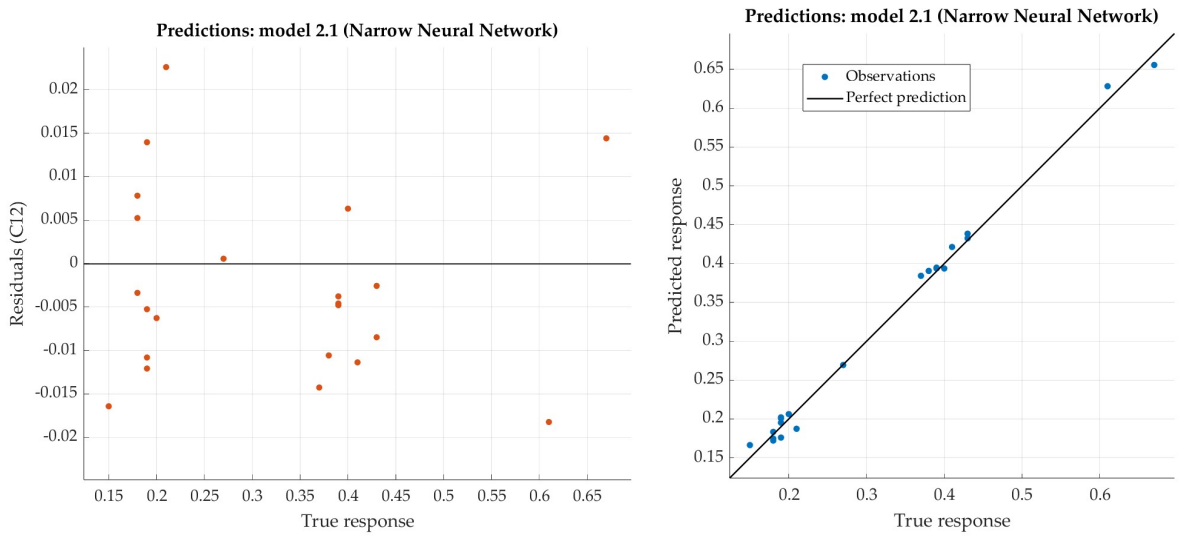


Figure 4.11. Performance evaluation of Narrow Neural Network on test dataset (carbon)

4.4.2. ANN model for manganese

The artificial neural network models for manganese are listed in Table 4.2, sorted by the lowest RMSE values for the test results.

Table 4.2. Results after training for manganese with six different architectures of ANNs

Model	Architecture	RMSE		R^2		MAE	MSE
		Train	Test	Train	Test	Test	Test
Narrow NN	26-10-1	0.1870	0.1122	0.485	0.823	0.0915	0.0126
Trilayered NN	26-10-10-10-1	0.1930	0.1163	0.451	0.810	0.0827	0.0135
Wide NN	26-100-1	0.2084	0.1251	0.361	0.780	0.0885	0.0157
Optimizable NN	26-3-1	0.1637	0.1362	0.605	0.739	0.1188	0.0186
Medium NN	26-25-1	0.2110	0.1665	0.344	0.610	0.1222	0.0277
Bilayered NN	26-10-10-1	0.2389	0.1918	0.160	0.482	0.1092	0.0368

Like the results for carbon, the Narrow Neural Network model has the best results. The predicted response vs. the experimental data, for each observation (from the training dataset), is shown in Figure 4.12. For the same model, the predicted response vs. the experimental unseen data (the test dataset) are shown in Figure 4.13. In Figure 4.13, the relation between residuals and real data are shown too. The best neural network model shows good results with relatively low RSME, as well as with satisfying R^2.

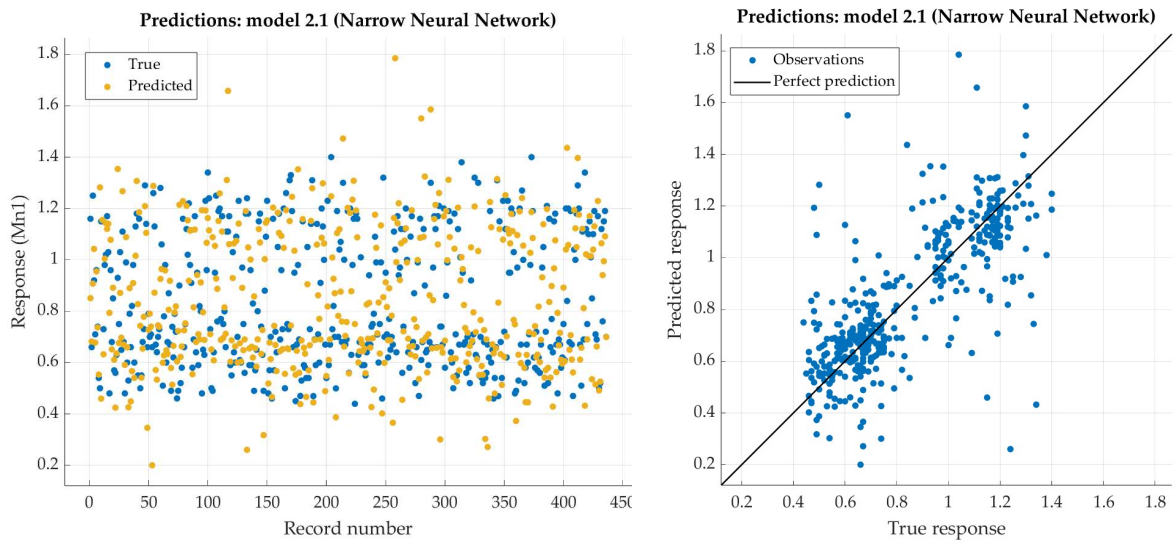


Figure 4.12. Performance evaluation of Narrow Neural Network on training dataset (manganese)

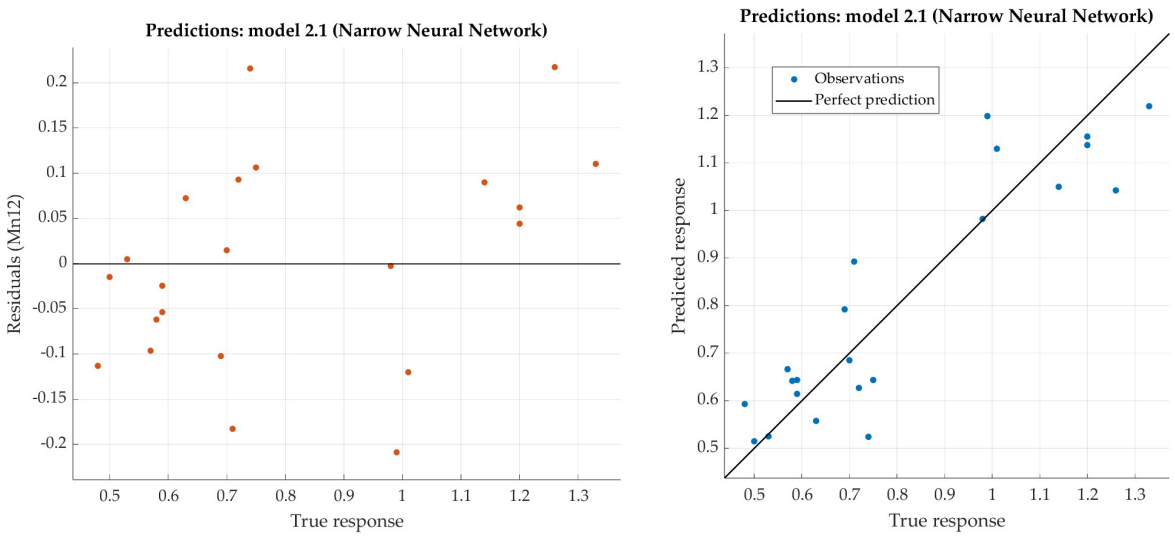


Figure 4.13. Performance evaluation of Narrow Neural Network on test dataset (manganese)

4.4.3. ANN model for silicon

The artificial neural network models for silicon are placed in Table 4.3 according to the lowest RMSE for test results as well. According to the given criterion, Optimizable Neural network is placed in first row. However, predicted values of this model are concentrated around the mean value of trained dataset. Due to that, the best model is chosen the Narrow Neural Network model. For that model, the predicted response vs. the experimental data, for each observation from the training dataset, is shown in Figure 4.14. For the same model, the predicted response vs. the experimental unseen data is shown in Figure 4.15. In Figure 4.15, the relation between residuals and true data are shown too.

The Narrow Neural Network has very low RSME, so one can conclude that the model is very good. At the same time, the coefficient of determination R^2 is negative. By definition, the coefficient of determination can be negative, indicating that the regression performed poorly—worse even than a model that explains none of the variability in the response data relative to its mean [85]. However, the residuals in Figure 4.15 are minimal, suggesting that the predicted values closely match the experimental data. Additionally, the predicted values for silicon fall within the specified steel standard.

Table 4.3. Results after training for silicon with six different architectures of ANNs

Model	Architecture	RMSE		R^2		MAE	MSE
		Train	Test	Train	Test	Test	Test
Optimizable NN	26-5-1	0.0426	0.0342	-0.011	-0.072	0.0258	0.0012
Narrow NN	26-10-1	0.0515	0.0371	-0.476	-0.262	0.0279	0.0014
Bilayered NN	26-10-10-1	0.0645	0.0401	-1.317	-0.474	0.0317	0.0016
Trilayered NN	26-10-10-10-1	0.0717	0.0428	-1.864	-0.678	0.0333	0.0018
Wide NN	26-100-1	0.0757	0.0510	-2.192	-1.387	0.0392	0.0026
Medium NN	26-25-1	0.0622	0.0630	-1.152	-2.642	0.0405	0.0040

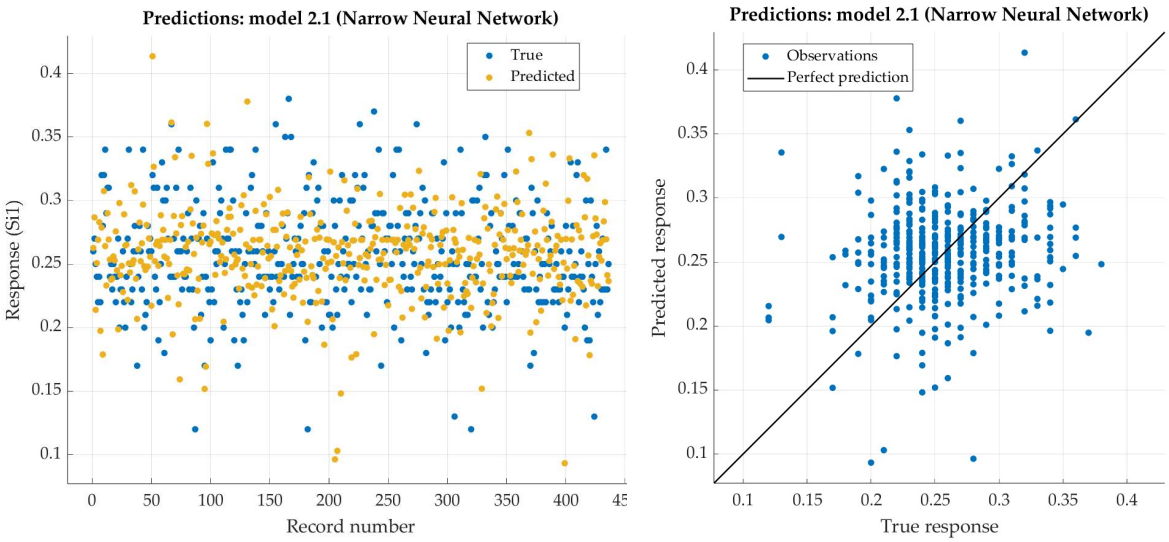


Figure 4.14. Performance evaluation of Narrow Neural Network on training dataset (silicon)

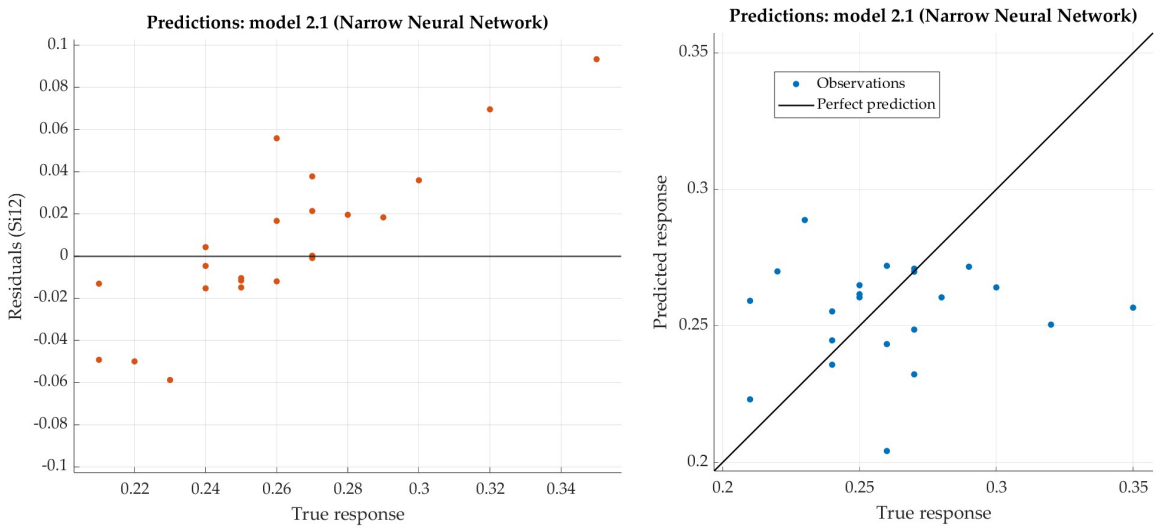


Figure 4.15. Performance evaluation of Narrow Neural Network on test dataset (silicon)

4.4.4. ANN model for chromium

The artificial neural network models for chromium are placed in Table 4.4 according to the lowest RMSE as well.

According to the given criterion, Bilayered Neural network model is chosen as the best. For that model, the predicted response vs. the experimental data, for each observation from the training dataset, is shown in Figure 4.16. For the same model, the predicted response vs. the experimental unseen data is shown in Figure 4.17. In Figure 4.17, the relation between residuals and real data is shown too.

The Bilayered Narrow Neural Network has relatively low RSME and very high the coefficient of determination R^2 (close to 0.9).

Table 4.4. Results after training for chromium with six different architectures of ANNs

Model	**Architecture**	**RMSE**		**R^2**		**MAE**	**MSE**
		Train	**Test**	**Train**	**Test**	**Test**	**Test**
Bilayered NN	26-10-10-1	0.1609	0.1168	0.748	0.896	0.1047	0.0824
Optimizable NN	26-219-1	0.1308	0.1377	0.834	0.856	0.0939	0.0886
Trilayered NN	26-10-10-10-1	0.1518	0.1400	0.776	0.851	0.0963	0.1009
Wide NN	26-100-1	0.1539	0.1447	0.770	0.841	0.1033	0.0839
Narrow NN	26-10-1	0.1433	0.1748	0.800	0.767	0.1006	0.1310
Medium NN	26-25-1	0.1814	0.2039	0.680	0.684	0.1196	0.1304

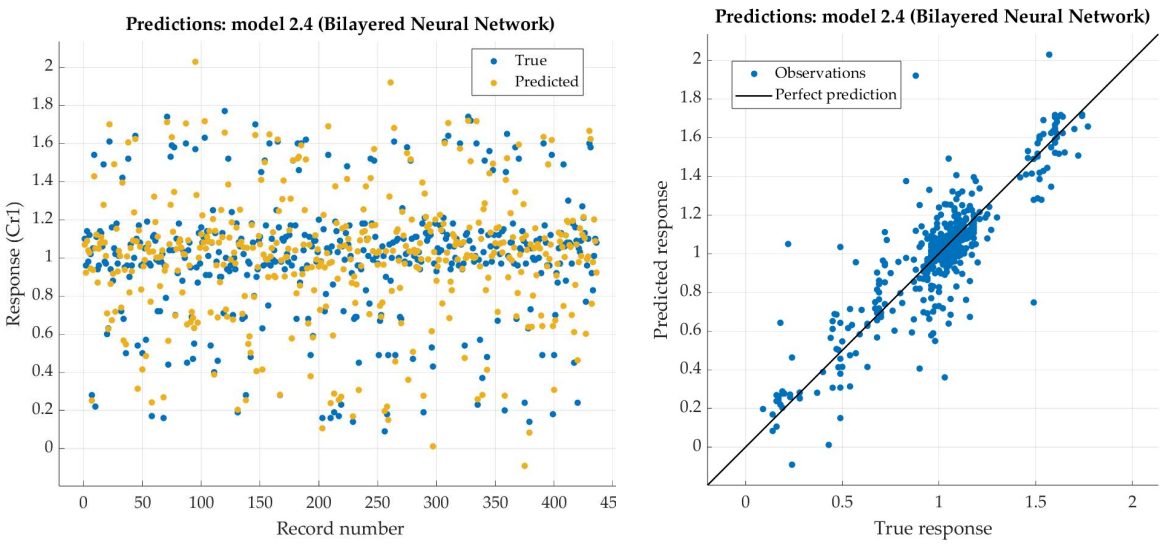


Figure 4.16. Performance evaluation of Bilayered Neural Network on training dataset (chromium)

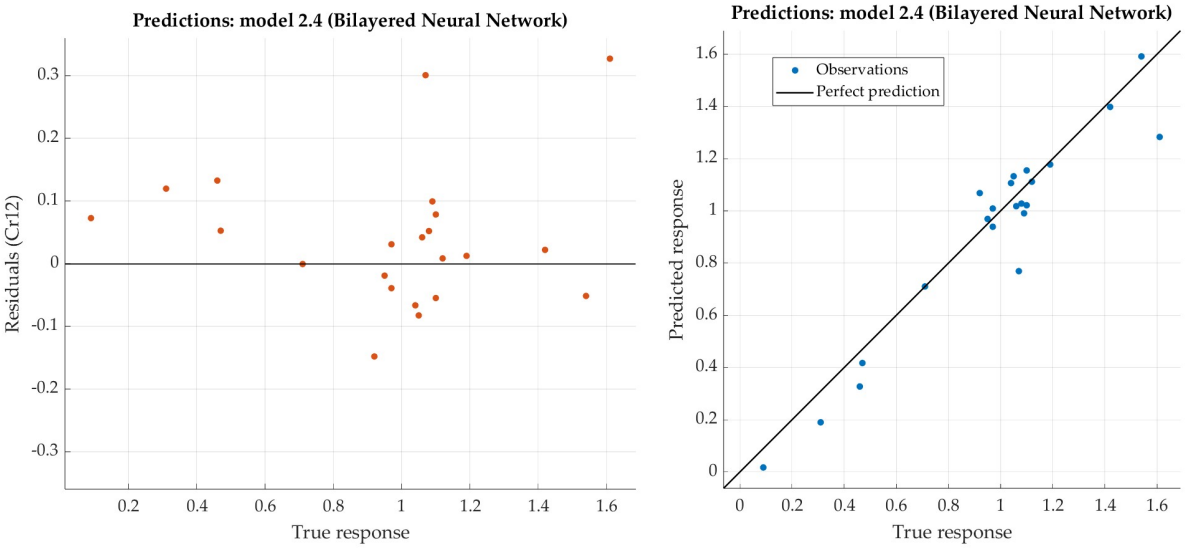


Figure 4.17. Performance evaluation of Bilayered Narrow Neural Network on test dataset (chromium)

4.4.5. ANN model for nickel

The artificial neural network models for nickel placed in Table 4.5 according to the lowest RMSE.

According to the given criterion, Trilayered Neural network model is chosen as the best (neural network with three hidden layers with 10 nodes each). For that model, the predicted response vs. the experimental data, for each observation from the training dataset, is shown in Figure 4.18. For the same model, the predicted response vs. the experimental unseen data is shown in Figure 4.19. In Figure 4.19 the relation between residuals and true data is shown too.

The Trilayered Narrow Neural Network has relatively low RSME and very high the coefficient of determination R^2 (close to 0.9). For most of the dataset, residuals are lower than 0.2 with one relatively high exception (0.6).

Table 4.5. Results after training for nickel with six different architectures of ANNs

Model	**Architecture**	**RMSE**		**R^2**		**MAE**	**MSE**
		Train	**Test**	**Train**	**Test**	**Test**	**Test**
Trilayered NN	26-10-10-10-1	0.2998	0.1590	0.660	0.899	0.0933	0.0253
Bilayered NN	26-10-10-1	0.3476	0.2386	0.544	0.771	0.1416	0.0569
Optimizable NN	26—1	0.2800	0.2450	0.704	0.759	0.1451	0.0600
Wide NN	26-100-1	0.3841	0.2715	0.443	0.704	0.1524	0.0737
Narrow NN	26-10-1	0.3537	0.3381	0.527	0.541	0.1986	0.1143
Medium NN	26-25-1	0.4303	0.3608	0.301	0.477	0.2241	0.1302

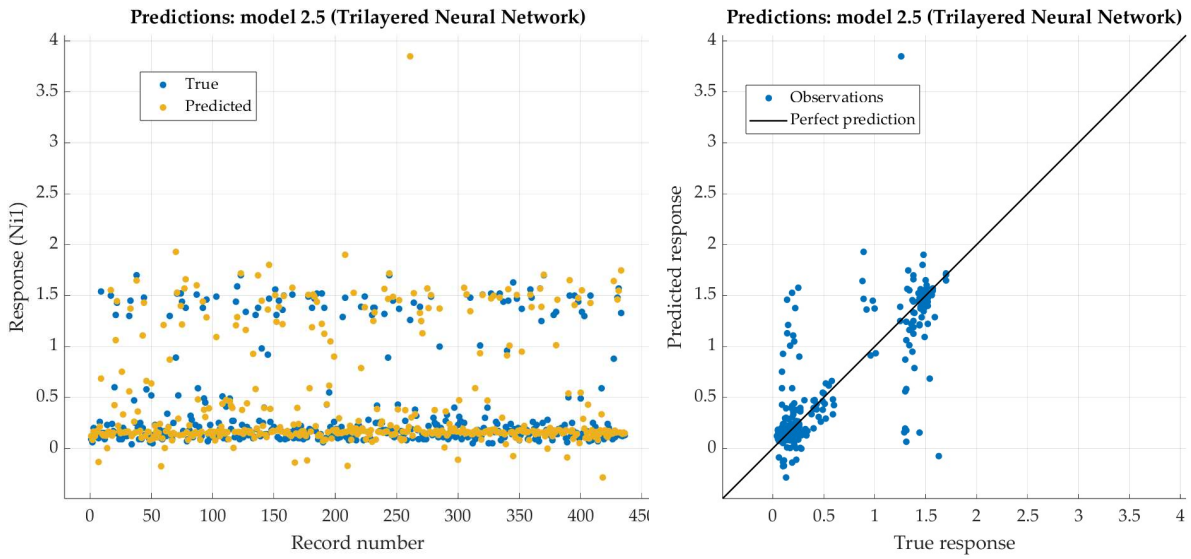


Figure 4.18. Performance evaluation of Trilayered Neural Network on training dataset (nickel)

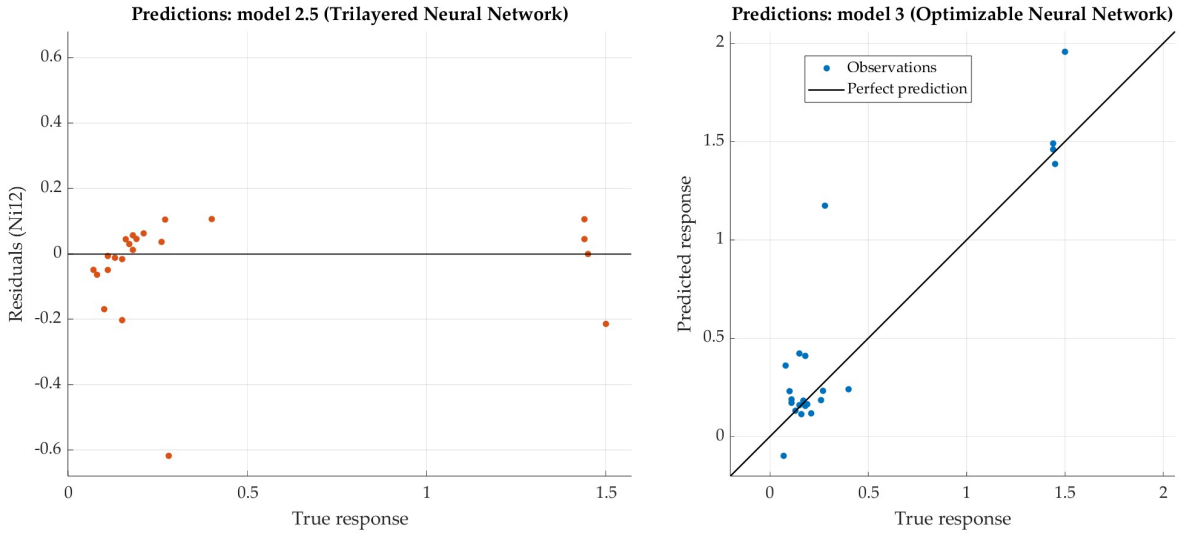


Figure 4.19. Performance evaluation of Trilayered Narrow Neural Network on test dataset (nickel)

4.4.6. ANN model for molybdenum

The artificial neural network models for molybdenum are placed in Table 4.6. according to the lowest RMSE as well.

According to the given criterion, Optimizable Neural network model is chosen as the best (neural network with two hidden layers, first with 295 nodes and second with 5 nodes). For that model, the predicted response vs. the experimental data, for each observation from the training dataset, is shown in Figure 4.20. For the same model, the predicted response vs. the experimental unseen data is shown in Figure 4.21. In Figure 4.21, the relation between residuals and true data is shown too.

The Optimizable Narrow Neural Network has very low RSME and weakly the coefficient of determination R^2. For most of the test dataset, residuals are lower than 0.1.

Table 4.6. Results after training for molybdenum with six different architectures of ANNs

Model	**Architecture**	**RMSE**		**R^2**		**MAE**	**MSE**
		Train	**Test**	**Train**	**Test**	**Test**	**Test**
Optimizable NN	26-295-5-1	0.0611	0.0570	0.5160	0.613	0.0442	0.0033
Trilayered NN	26-10-10-10-1	0.0871	0.0634	0.0147	0.521	0.0417	0.0040
Bilayered NN	26-10-10-1	0.2185	0.0682	-5.1983	0.445	0.0461	0.0047
Narrow NN	26-10-1	0.0798	0.0701	0.1732	0.415	0.0591	0.0049
Wide NN	26-100-1	0.1012	0.0832	-0.3292	0.176	0.0558	0.0069
Medium NN	26-25-1	0.1009	0.1070	-0.3205	-0.364	0.0787	0.0115

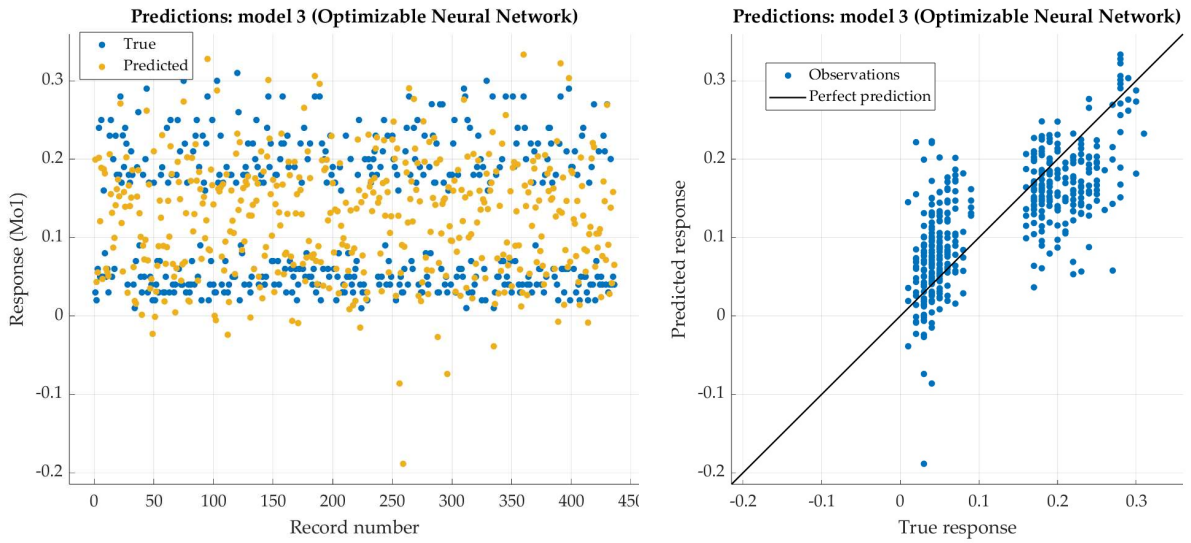


Figure 4.20. Performance evaluation of Optimized Neural Network on training dataset (molybdenum)

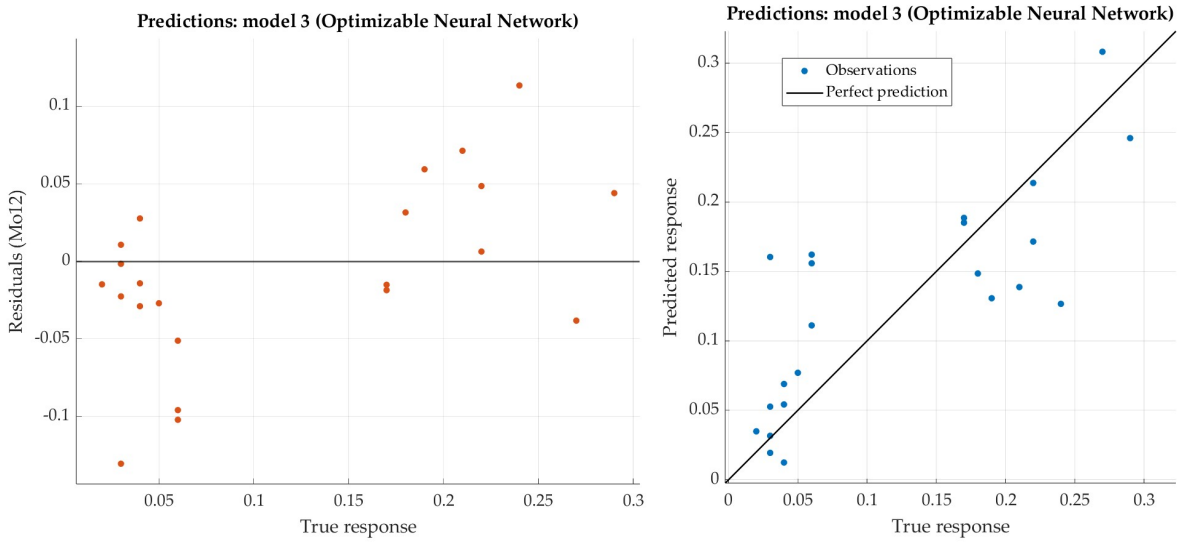


Figure 4.21. Performance evaluation of Optimized Narrow Neural Network on test dataset (molybdenum)

4.4.7. ANN model for copper

The artificial neural network models for copper are placed in Table 4.7 according to the lowest RMSE as well.

According to the given criterion (lowest RMSE), Optimizable Neural network is placed in first row (one hidden layer with 260 nodes). For that model, the predicted response vs. the experimental data, for each observation from the training dataset, is shown in Figure 4.22. For the same model, the predicted response vs. the experimental unseen data is shown in Figure 4.23. In Figure 4.23, the relation between residuals and true data are shown too.

The Optimizable Neural Network has very low RSME, while the coefficient of determination R^2 negative which shows that regression performed poorly. Nevertheless, residuals in figure 17 show very small figures (lower than 0.1). It means that predicted values are similar to the experimental data.

Table 4.7. Results after training for copper with six different architectures of ANNs

Model	**Architecture**	**RMSE**		**R^2**		**MAE**	**MSE**
		Train	**Test**	**Train**	**Test**	**Test**	**Test**
Optimizable NN	26-260-1	0.0442	0.0020	0.1505	0.0331	0.0297	0.0013
Narrow NN	26-10-1	0.0613	0.0038	-0.6341	0.0452	0.0433	0.0036
Medium NN	26-25-1	0.0659	0.0043	-0.8875	0.0462	0.0443	0.0042
Trilayered NN	26-10-10-10-1	0.0536	0.0029	-0.2493	0.0394	0.0516	0.0049
Wide NN	26-100-1	0.0741	0.0055	-1.3885	0.0500	0.0562	0.0052
Bilayered NN	26-10-10-1	0.0738	0.0055	-1.3784	0.0457	0.0639	0.0098

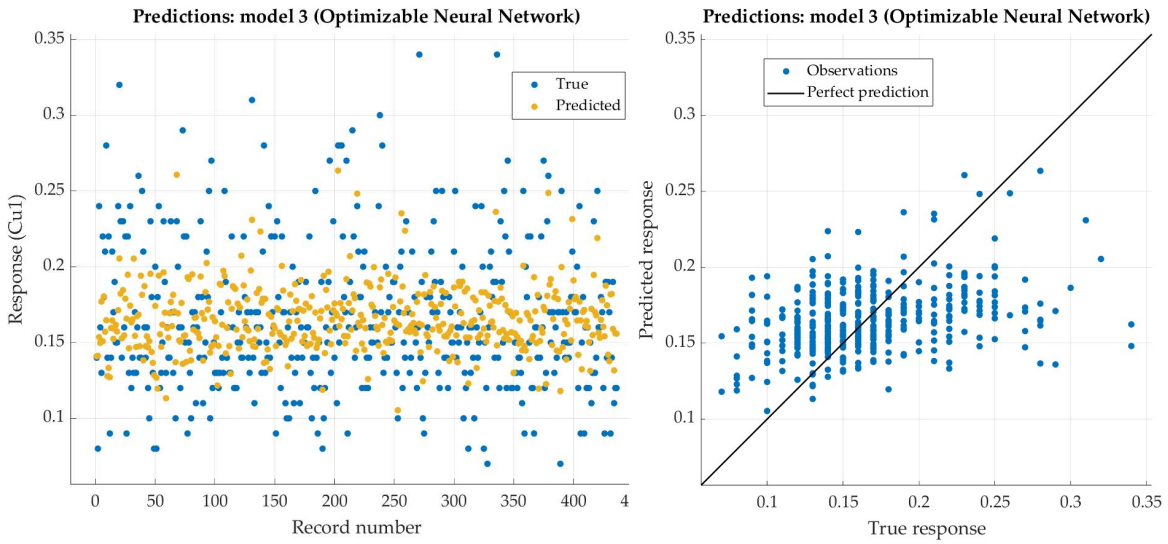


Figure 4.22. Performance evaluation of Narrow Neural Network on training dataset (copper)

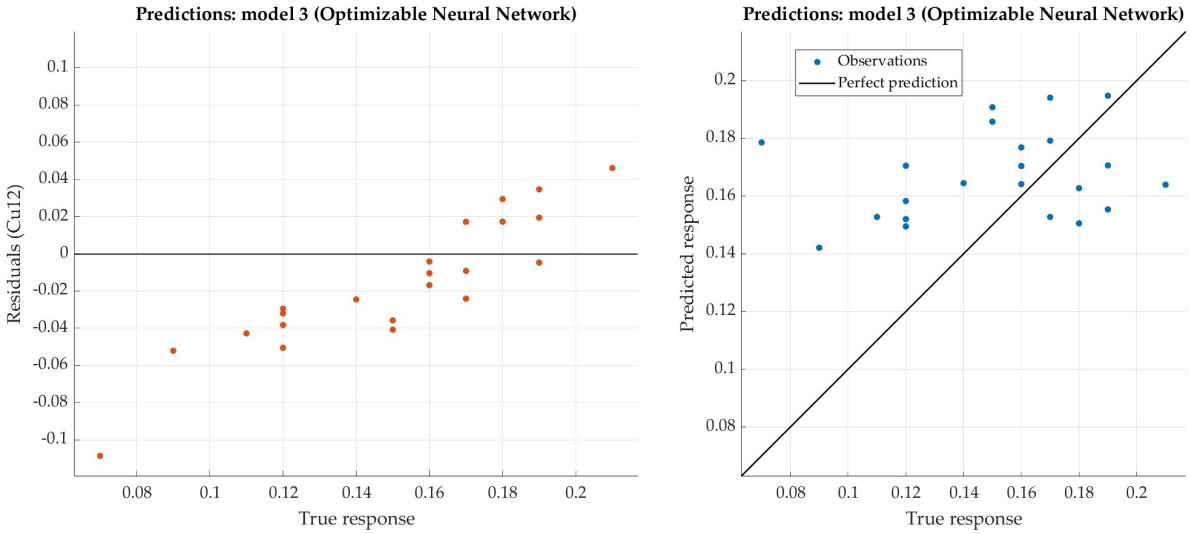


Figure 4.23. Performance evaluation of Optimized Narrow Neural Network on test dataset (copper)

5. EXPERIMENTAL VERIFICATION

To experimentally validate the model, five steels with varying hardenability were used: C60E, 41Cr4, 46Cr2, 17CrNi6-6, and 65Mn4. A comparison of their Jominy curves is shown in Figure 5.1.

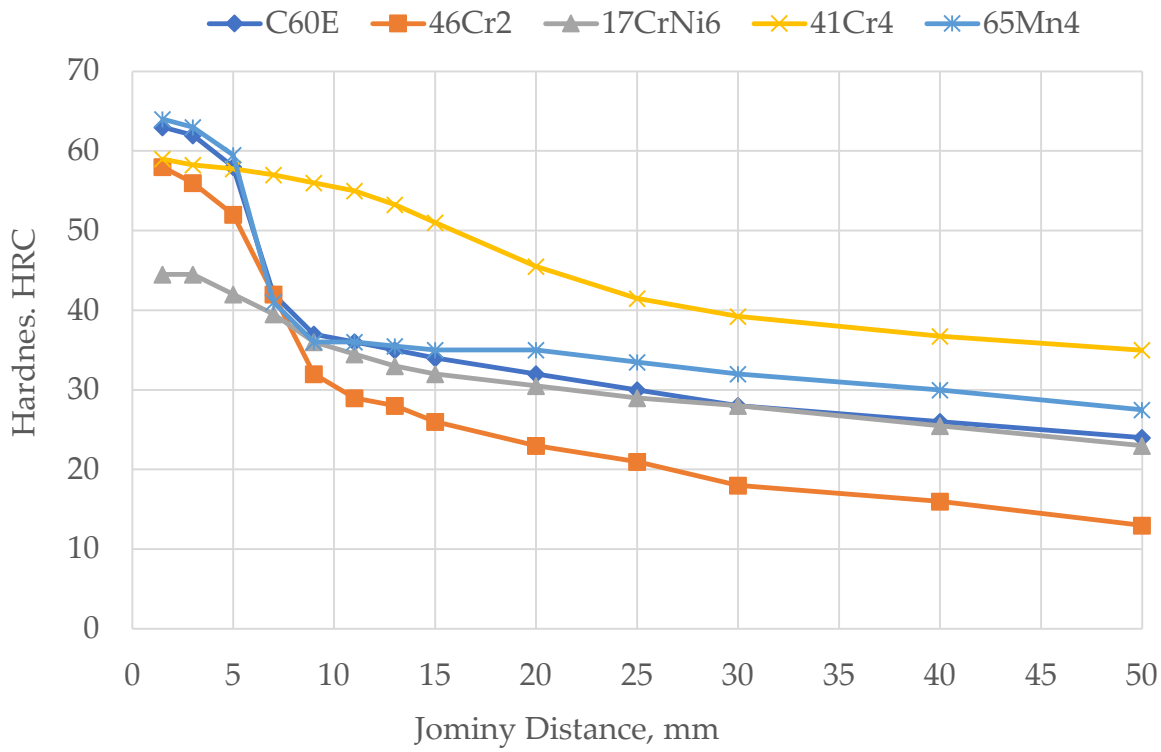


Figure 5.1. Jominy curves for five steels with different hardenability

Steel C60E is chosen from the EN10083-2:2006 standard (non-alloy steels for quenching and tempering) as a non-alloy steel. Steel C60E can be defined as steel with low hardenability.

Alloy steels 41Cr4 and 46Cr2 are steel grades chosen from the standard EN10083-3:2007 (alloy steels for quenching and tempering). Hardenability of the steel 41Cr4 can be classified as moderate to high. Steel 46Cr2 can be considered as a steel with low to moderate hardenability.

Steel 17CrNi6-6 is the steel grade chosen from standard EN10084:2008 (case-hardening steels). The steel 17CrNi6-6 has a low carbon content, hence the maximal hardness is 50 HRC. A relatively high content of chromium and nickel makes the steel capable of achieving high hardness at significant depths and thus can be considered as a steel with high hardenability.

65Mn4 is DIN Steel number (1.1240) grade. This steel exhibits relatively high hardness values near the quenched end of the Jominy specimen. However, the hardness declines significantly after a distance of only 7 mm from the quenched end, which characterizes it as having low hardenability.

Jominy curves for five steels are shown in Figures 5.2 to 5.6, while their chemical compositions are presented in Tables 5.1 to 5.5.

According to the given Jominy curves of five steels, chemical compositions are predicted using ANN models for Approach 2 to Approach 4 of Dataset organization. Predicted values are compared to the real data.

The Jominy curve of the steel C60E is shown in Figure 5.2.

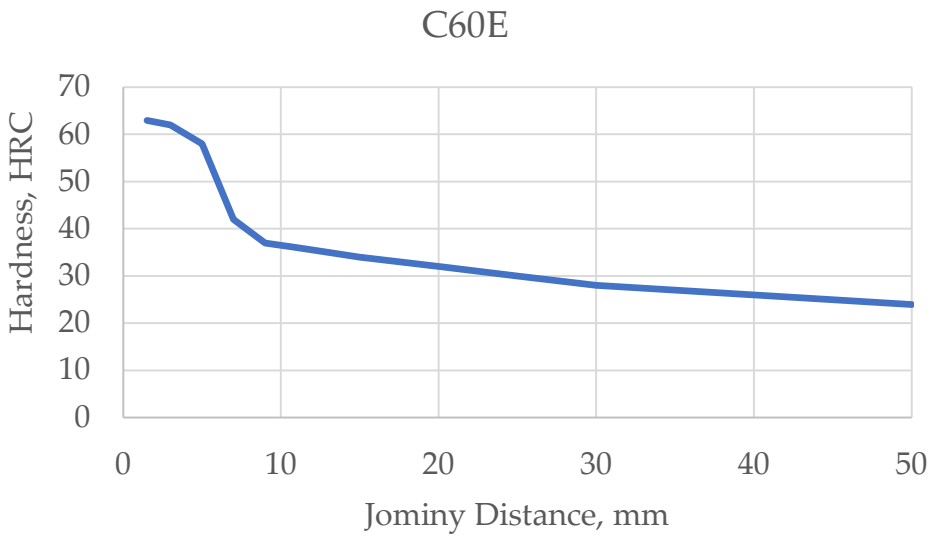


Figure 5.2. Jominy curve for the steel grade C60E

Chemical composition of the steel grade C60E is given in Table 5.1.

Table 5.1. Chemical composition of the steel C60E

C	Mn	Si	Cr	Ni	Mo	Cu
0.61	0.69	0.24	0.31	0.11	0.03	0.19

The Jominy curve of the steel 41Cr4 is shown in Figure 5.3.

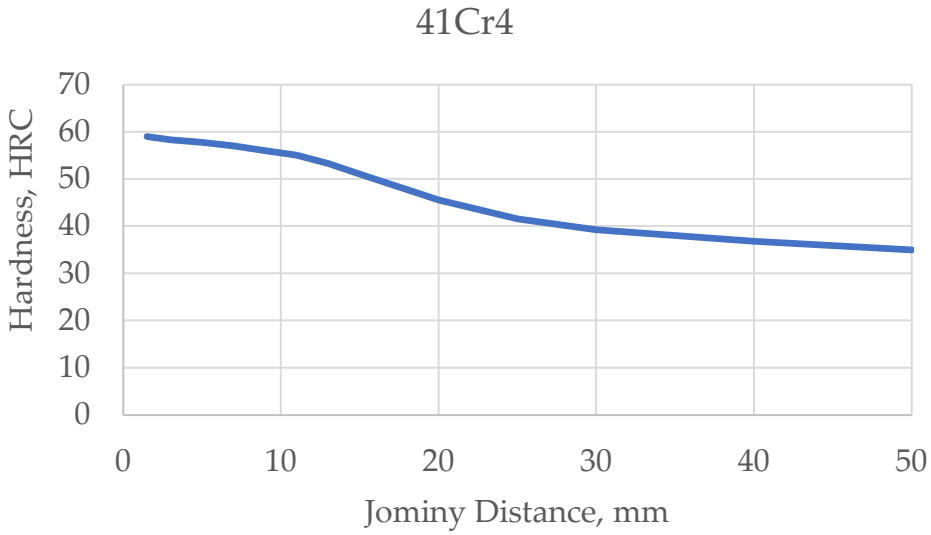


Figure 5.3. Jominy curve for the steel grade 41Cr4

Chemical composition of the steel grade 41Cr4 is given in Table 5.2.

Table 5.2. Chemical composition of the steel 41Cr4

C	Mn	Si	Cr	Ni	Mo	Cu
0.43	0.72	0.27	1.09	0.15	0.06	0.16

The Jominy curve of the steel 46Cr2 is shown in Figure 5.4.

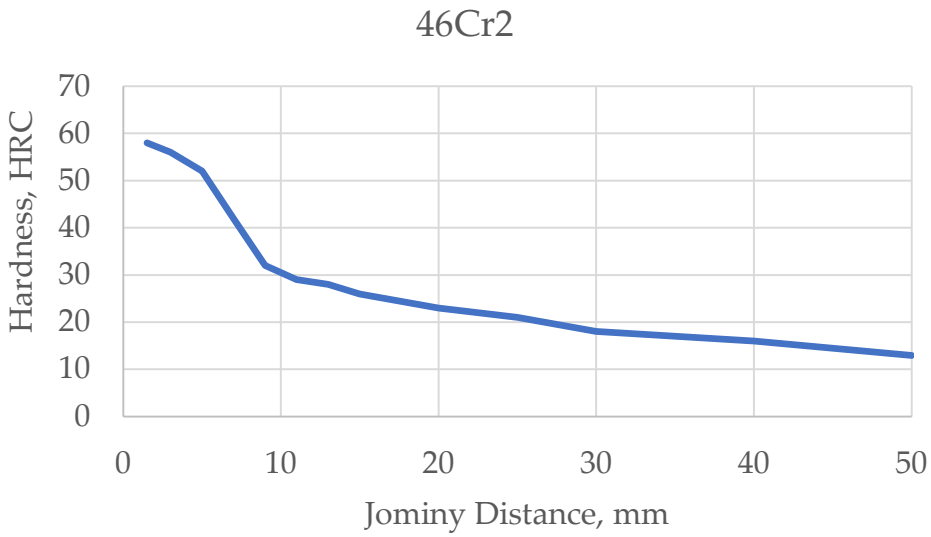


Figure 5.4. Jominy curve for the steel grade 46Cr2

Chemical composition of the steel grade 46Cr2 is given in Table 5.3.

Table 5.3. Chemical composition of the steel 46Cr2

C	Mn	Si	Cr	Ni	Mo	Cu
0.43	0.59	0.26	0.47	0.40	0.04	0.17

The Jominy curve of the steel 17CrNi6-6 is shown in Figure 5.5.

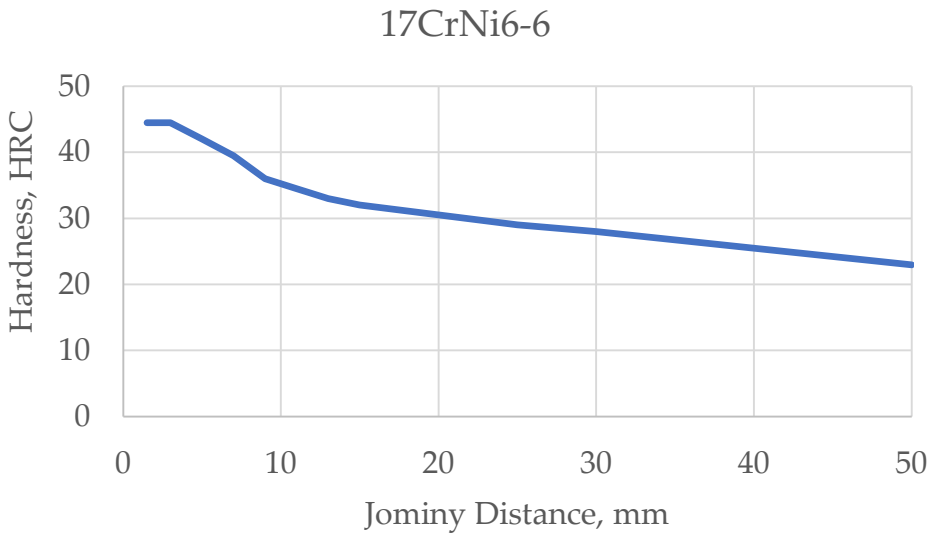


Figure 5.5. Jominy curve for the steel grade 17CrNi6-6

Chemical composition of the steel grade 17CrNi6-6 is given in Table 5.4.

Table 5.4. Chemical composition of the steel 17CrNi6-6

C	Mn	Si	Cr	Ni	Mo	Cu
0.15	0.53	0.22	1.42	1.45	0.04	0.17

The Jominy curve of the steel 65Mn4 is shown in Figure 5.6.

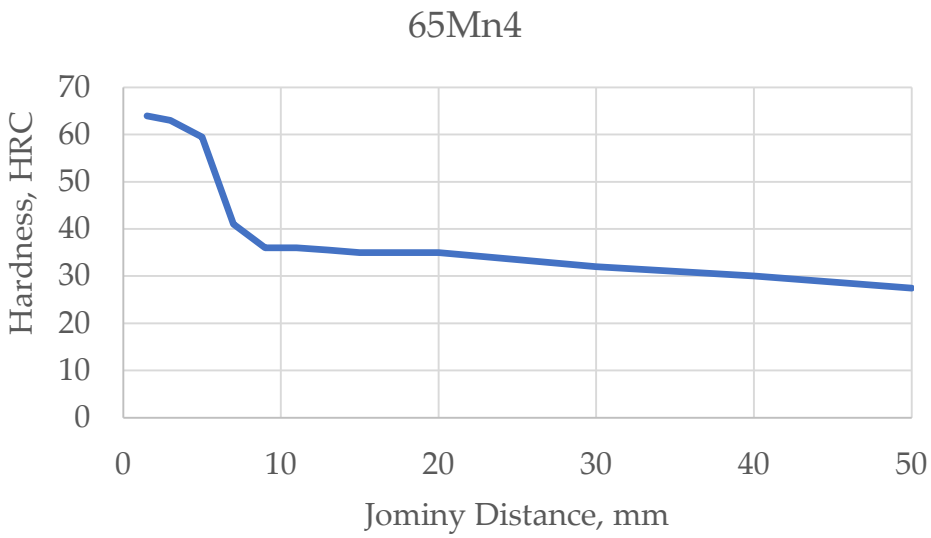


Figure 5.6. Jominy curve for the steel grade 65Mn4

Chemical composition of the steel grade 65Mn4is given in Table 5.5.

Table 5.5. Chemical composition of the steel 65Mn4

C	Mn	Si	Cr	Ni	Mo	Cu
0.67	1.01	0.32	0.09	0.07	0.02	0.15

The predicted values of alloying elements for different approaches of data representation, for each of five steels (C60E, 41Cr4, 46Cr2, 17CrNi6-6 and 65Mn4) are presented in separate Tables 5.6 to 5.10, respectively. For each of these steels, the minimum and maximum values of the defined chemical elements, according to the steel grades, are presented in Tables 5.6 to 5.10.

Table 5.6. Experimental vs. predicted data of chemical elements for steel C60E

	C	**Mn**	**Si**	**Cr**	**Ni**	**Mo**	**Cu**
Predicted values – Approach 1	Calculation has not been made due to discontinuing						
Predicted values – Approach 2	0.56	0.72	0.25	0.29	0.05	0.03	0.19
Predicted values – Approach 3	0.61	0.73	0.27	0.30	0.11	0.01	0.18
Predicted values – Approach4	0.63	0.79	0.26	0.19	0.12	0.02	0.19
Experimental values	0.61	0.69	0.24	0.31	0.11	0.03	0.19
Steel grade limits	0.57 – 0.65	0.6 – 0.9	< 0,4	< 0,4	< 0,4	< 0,1	N/D*

* N/D not defined limits within the steel grade

Table 5.7. Experimental vs. predicted data of chemical elements for steel 41Cr4

	C	**Mn**	**Si**	**Cr**	**Ni**	**Mo**	**Cu**
Predicted values – Approach 1	Calculation has not been made due to discontinuing						
Predicted values – Approach 2	0.43	0.66	0.26	1.04	0.09	0.08	0.15
Predicted values – Approach 3	0.41	0.69	0.27	1.04	0.12	0.07	0.14
Predicted values – Approach4	0.42	0.69	0.26	1.03	0.14	0.07	0.15
Experimental values	0.41	0.7	0.25	1.08	0.13	0.04	0.12
Steel grade limits	0.38 – 0.45	0.6 – 0.9	< 0,4	0.9–1.2	N/D*	N/D*	N/D*

* N/D not defined limits within the steel grade

Table 5.8. Experimental vs. predicted data of chemical elements for steel 46Cr2

	C	**Mn**	**Si**	**Cr**	**Ni**	**Mo**	**Cu**
Predicted values – Approach 1	Calculation has not been made due to discontinuing						
Predicted values – Approach 2	0.41	0.62	0.24	0.46	0.18	0.10	0.21
Predicted values – Approach 3	0.44	0.62	0.25	0.49	0.36	0.03	0.18
Predicted values – Approach4	0.44	0.61	0.20	0.42	0.29	0.01	0.20
Experimental values	0.43	0.59	0.26	0.47	0.40	0.04	0.17
Steel grade limits	0.42 – 0.50	0.5 – 0.8	< 0,4	0.4–0.6	N/D*	N/D*	N/D*

* N/D not defined limits within the steel grade

Table 5.9. Experimental vs. predicted data of chemical elements for steel 17CrNi6-6

	C	**Mn**	**Si**	**Cr**	**Ni**	**Mo**	**Cu**
Predicted values – Approach 1	Calculation has not been made due to discontinuing						
Predicted values – Approach 2	0.15	0.64	0.23	1.50	1.28	0.06	0.16
Predicted values – Approach 3	0.18	0.58	0.24	1.42	1.37	0.01	0.17
Predicted values – Approach4	0.17	0.53	0.27	1.40	1.45	0.05	0.15
Experimental values	0.15	0.53	0.22	1.42	1.45	0.04	0.17
Steel grade limits	0.14 – 0.20	0.5 – 0.8	< 0,4	1.4–1.7	1.4–1.7	N/D*	N/D*

* N/D not defined limits within the steel grade

Table 5.10. Experimental vs. predicted data of chemical elements for steel 65Mn4

	C	**Mn**	**Si**	**Cr**	**Ni**	**Mo**	**Cu**
Predicted values – Approach 1	Calculation has not been made due to discontinuing						
Predicted values – Approach 2	0.60	0.90	0.26	0.09	0.05	0.03	0.19
Predicted values – Approach 3	0.66	0.97	0.28	0.09	0.05	0.01	0.17
Predicted values – Approach4	0.66	1.13	0.25	0.02	0.12	0.03	0.19
Experimental values	0.67	1.01	0.32	0.09	0.07	0.02	0.15
Steel grade limits	0.6 – 0.7	0.9 – 1.2	0.25 – 5	N/D*	N/D*	N/D*	N/D*

* N/D not defined limits within the steel grade

Predicted values for chemical elements, for all steels, are within the limits defined by the steel grades. The model was validated accordingly, confirming its accuracy.

6. DISCUSSION OF RESULTS

Contemporary industry demands, particularly in materials and steel production, focus on achieving specific hardenability requirements. Along with these demands, new insights have emerged into the influence of different alloying elements and heat treatment processes on hardenability.

Customers are often required to choose a steel grade that provides the desired hardenability. However, this steel grade may have excessive hardenability, which unnecessarily increases the use of alloying elements. Therefore, it is important to design the chemical composition according to the customer's requirements in order to achieve the desired hardenability while keeping production costs relatively low. The demand for specific hardenability typically falls within a narrow range that is generally covered by a specific steel grade. Sometimes, customer demands are not only restricted by the lower hardenability limit but also by the upper hardenability limit.

This research has led to new possibilities for better predicting and controlling hardenability. However, despite these developments, defining the precise relationships between key factors—such as chemical composition and microstructure—and the resulting hardenability remains a complex and challenging task. The current state of the art in the field does not provide enough knowledge to establish a mathematical relationship or model between chemical composition and the hardenability of steel. In support of this thesis, existing models are not sufficiently precise or are limited to specific groups of steels [54, 55, 60].

On the other hand, calculating the chemical composition based on the required Jominy curve shape (which is inverse problem) provides a new approach to this challenge. This task can be tackled using contemporary computational methods, particularly neural networks. An artificial neural network acts as a black-box model capable of analyzing previously undiscovered phenomena during phase transformations in heat

treatment, as well as the synergistic influence of alloying elements on properties, including hardenability. Due to these facts in this thesis was proposed the methodology of designing chemical composition based on the required hardenability. A representative dataset of 470 steel samples was collected, with each sample including Jominy test results and chemical compositions. Additionally, an indication of significant martensite presence in the microstructure (with the limit set at 50%) was included in the input data. The input features consisted of Jominy test results and indicators of martensite presence, while the output target consisted of the mass fractions of seven alloying elements (C, Mn, Si, Cr, Ni, Mo and Cu). Various neural network models were developed for different dataset representations. The models were compared based on their predicted results and experimentally validated using a selected steel grade with known chemical composition and significantly different Jominy curve.

The key factor determining the success of a neural network architecture is the representation of the dataset. MATLAB 2023b was used to develop the different neural network architectures. In this doctoral thesis, four distinct approaches to dataset representation were investigated (Sections 3.3 to 3.6). Each new approach offered better results and more options for future research. In Approach 1, the input features were limited to only 14 parameters: 13 hardness values at specific distances from the quenched end and one indicating the presence of martensite. An attempt was made to associate the presence of martensite with a specific distance from the quenched end. The hardness value for that distance was included, while hardness values for other distances were set to zero. With this approach, 13 observations were created for each steel sample (Table 3.7). The dataset was randomly divided into three subsets: the training dataset (70% of the total data), the validation dataset (15% of the total data), and the test dataset (15% of the total data). For this dataset representation, a neural network with the architecture denoted as [14-27-27-7] was applied. This architecture indicates that the input layer consisted of 14 nodes, two hidden layers each containing

27 nodes, and an output layer with 7 nodes corresponding to 7 chemical elements. Unfortunately the obtained results were physically meaningless, suggesting that the model was not learning effectively from the training data (Figures 4.1 to 4.3). Therefore, further research on this model was discontinued.

In Approach 2, the idea was that one pattern is related to one steel. To more clearly link the hardness values and the indications of martensite presence at specific distances from the quenched end, the input pattern consisted of 39 data points. The input features were consisted of 39 parameters: 13 Jominy distances from quenched end of Jominy specimen (J_d), 13 hardness values at specific distances from the quenched end and 13 indicators of the presence of martensite in the microstructure (Table 3.9). Outputs (responses) were mass fractions of seven alloying elements. The dataset was randomly divided into three subsets: the training dataset (70% of the total data), the validation dataset (15% of the total data), and the test dataset (15% of the total data). For this dataset representation, among others a neural network with the architecture denoted as [39-14-14-7] was applied. This architecture indicates that the input layer consisted of 39 nodes, two hidden layers each containing 14 nodes, and an output layer with 7 nodes corresponding to 7 alloying elements. The results are presented in Figures 4.4. to 4.6. For Approach 2 of data representation and the proposed neural network architecture, the following performance metrics were obtained: RMSE = 0.121 and the coefficient of determination $R^2 = 0.88$.

The next optimization of the dataset representation focused on reducing the number of input parameters. Distances from the quenched end were removed, so the input features consisted of 26 parameters: 13 hardness values at specific distances from the quenched end and 13 indicators of the presence of martensite in the microstructure (Table 3.10). The dataset was randomly divided into three subsets: the training dataset (70% of the total data), the validation dataset (15% of the total data), and the test dataset (15% of the total data). For Approach 3 of dataset representation, among others a neural network with the architecture denoted as [26-14-14-7] was applied. For

Approach 3 of data representation and the proposed neural network architecture, the following performance metrics were obtained: RMSE=0.123 and the coefficient of determination $R^2 = 0.90$ (Figures 4.7 to 4.9).

It is important to mention, that some predicted chemical compositions obtained using Approaches 2 and 3 were not within the limits of the specified steel grades. Therefore, in order to solve any potential problems in situations where non-standard chemical compositions are obtained as a result of calculations, it was decided to organize the data set in a different way – Approach 4, one that would allow correction of the concentration of an element whose calculated concentration was outside the allowable limits for a particular steel grade

Approach 4 of data representation focused on modeling each alloying element separately. The input features remained the same as in Approach 3, consisting of 26 parameters: 13 hardness values at specific distances from the quenched end and 13 indicators of the presence of martensite in the microstructure (shown in Table 3.11). The output consisted of one parameter (one alloying element). Fo Approach 4 of dataset representation the Machine Learning and Deep Learning Toolbox from MATLAB R2023b was used for regression tasks. Neural network regression models are trained using the Regression Learner App, with 10-fold cross-validation. Five distinct artificial neural network architectures were used for training. All neural networks had an input layer with 26 nodes and an output layer with one node. The hidden layers varied across the different architectures. The narrow network consisted of one hidden layer with 10 nodes. The medium network had one hidden layer with 25 nodes. The broad network included one hidden layer with 100 nodes. The bilayered network had two hidden layers, each containing 10 nodes. The trilayered network featured three hidden layers, each with 10 nodes. The artificial neural network models were sorted according to their lowest RMSE values for the test results in the Tables 4.1 to 4.7.

If the predicted mass fraction of an alloying element exceeds the limits of the steel grade, a specific adjustment of the mass fraction is possible for that particular alloying element. The accuracy of the neural network models is particularly high for carbon, manganese, chromium, and nickel, as indicated by low RMSE values and high R^2 values (Table 4.1, Table, 4.2, Table, 4.4 and Table 4.5). This is particularly important because these elements have a strong influence on hardenability. The mass fraction of silicon exhibits significant variability in the dataset, with values that are practically random. This randomness reduced the accuracy of the neural network model, leading to higher RMSE values and lower R^2 values (Table 4.3). Silicon remains in steel after the metallurgical process, primarily as a residual deoxidizer. While its concentration can vary due to processing factors and is essentially random, it typically remains within controlled limits. The residuals, which represent the difference between predicted and real experimental values for copper, are very small. Although the coefficient of determination for copper is low (Table 4.7), similar to silicon, the predicted values are very close to the experimental values (Figure 4.17). The coefficient of determination for molybdenum is higher, and the predicted values are also very close to the experimental values (Figure 4.15).

For the experimental verification of the neural network models, five steel with distinct hardenability has chosen (Figure 5.1). The predicted chemical compositions obtained by neural network models for different approaches to dataset representation for five representative steels (C60E, 41Cr2, 46Cr2, 17CrNi6 and 65Mn4) were presented in Tables 5.6 to 5.10 include the experimental values for these steels and the corresponding steel grade limits. All neural network models provided predicted chemical compositions that fall within the specified steel grade limits.

The proposed methodology for designing the chemical composition of steel based on the required hardenability can offer significant benefits to steel producers. This approach ensures that the predicted chemical composition results in a very narrow range of the desired hardenability (Jominy curve). This approach is innovative for the

automated and precise prediction of a chemical composition of steel. To design any artificial neural network a relatively large number of experimental data are required. With additional experimental data, for nickel and molybdenum steels especially, used for learning and testing neural networks, the proposed models can be further tuned and improved.

7. CONCLUSION

This thesis introduces an innovative approach for the automated and precise prediction of the chemical composition of steel based on the Jominy curve, which represents the required hardenability. The method also accounts for microstructure (martensite presence) at different distances from the quenched end of the Jominy specimen.

Unlike traditional hardenability modelling methods, which determine the Jominy curve for a given chemical composition, this approach makes possible to solve inverse problem, designing the chemical composition from the required Jominy curve. Solving of the inverse problem in this research is the new and innovative contribution in the field of material science.

The research results obtained in this work allow to formulate the following conclusions:

1. Using neural networks, models of the relationship between hardenability and chemical composition of steel was developed taking into account the microstructure (martensite presence), based on an appropriate data set, including successive hardness values on the Jominy curve, together with the mass fraction of alloying elements C, Si, Mn, Cr, Ni, Mo and Cu.
2. The developed neural network models yielded acceptable results for designing the chemical composition of steel. These predicted results were experimentally validated with four representative steel grades (C60E, 41Cr4, 46Cr2, and 17CrNi6-6) from three different standards (EN 10083-2, EN 10083-3, and EN 10084), as well as with the additional steel grade 65Mn4.
3. The developed artificial neural network models demonstrate their effectiveness in predicting the chemical composition of steels for heat treatment based on the curvature of the Jominy curve. The predicted values for all seven chemical

elements align closely with the experimental values and fall within the limits defined by the respective steel grades.

4. This methodology for selecting the optimal steel composition based on the required hardenability has significant practical applications in the mechanical engineering and manufacturing sectors. By using this approach, chemical compositions can be designed to meet specific customer requirements while ensuring the desired hardenability. Furthermore, the ability to achieve these results at relatively low production costs makes this method both efficient and economically viable, providing a valuable tool for steel design and selection in industrial applications.

BIBLIOGRAPHY

[1] Stupnišek M, Cajner F. *Osnove toplinske obradbe metala*. Zagreb (HR): Sveučilište u Zagrebu; 2001. Croatian.

[2] Brooks CR. *Principles of the Heat Treatment of Plain Carbon and Low Alloy Steels*. Materials Park (OH): ASM International; 1996.

[3] Dossett JL. Structure of metals and alloys. In: *Practical Heat Treating*. Materials Park (OH): ASM International; 2020. p. 1–27.

[4] Smoljan B. *Mikrostrukturne pretvorbe, nosivosti, načela toplinske obrade metala*. Varaždin (HR): Sveučilište Sjever; 2020. Croatian.

[5] Bhadeshia HKDH. Physical metallurgy of steel. In: *Physical Metallurgy*. 5th ed. 2014. p. 2157–2214.

[6] Canale LCF, Vatavuk J, Totten GE. Introduction to steel heat treatment. In: *Comprehensive Materials Processing*. Vol. 12. 2014. p. 1–38.

[7] Krauss G. Martensite before and after tempering, deformation and fracture. In: *Encyclopedia of Materials: Science and Technology*. 2016. p. 5193–7.

[8] Maki T, Tsuzaki K, Tamura I. The morphology of microstructure composed of lath martensite. *Trans ISIJ*. 1980;20:207–14.

[9] Cai J, Lin J. Modelling phase transformations in hot stamping and cold die quenching of steels. In: *Microstructure Evolution in Metal Forming Processes*. 2012. p. 210–36.

[10] Takahashi M. Recent progress: kinetics of the bainite transformation in steels. *Curr Opin Solid State Mater Sci*. 2004;8(3–4):213–17.

[11] Liščić B. Hardenability. In: Totten GE, editor. *Steel Heat Treatment Handbook*. 2nd ed. Boca Raton (FL): CRC Press; 2007. p. 213–76.

[12] Edenhofer B, Joritz D, Rink M, Voges K. Carburizing of steels. In: *Thermochemical Surface Engineering of Steels*. 2015. p. 485–553.

[13] Verhoeven JD, editor. Hardenability of steel. In: *Steel Metallurgy for Non-Metallurgist*. Materials Park (OH): ASM International; 2007. Chapter 9. DOI: 10.1361/smnm2007p083.

[14] Bhargava AK, Banerjee MK. Hardenability of steel. In: *Comprehensive Materials Finishing*. Vol. 2. Surface and Heat Treatment Processes. 2017. p. 55–70.

[15] Jominy WE, Boegehold AL. A hardenability test for carburising steel. *Trans ASM*. 1938;26:574–606.

[16] Inoue T. Metallo-thermo-mechanical coupling in quenching. In: Hashmi S, Batalha GF, editors. *Comprehensive Materials Processing*. Vol. 12. Amsterdam (NL): Elsevier; 2014. p. 177–251.

[17] Landek D, Liščić B, Filetin T, Looben T. End-quench hardenability test for gas quenched steels. *Strojarstvo*. 2011;53(1):33–7.

[18] Canale LCF, Albano LL. Hardenability of steel. In: *Comprehensive Materials Processing*. Vol. 12. 2014. p. 39–97.

[19] Chi C, Pian L, Gu J, Sun Y. Effect of Cr on the microstructure and strength–toughness of high-strength and heat-resistant stainless steel. *Steel Res Int*. 2024;95:2400412.

[20] Wang LJ, Cai QW, Wu HB, et al. Effects of Si on the stability of retained austenite and temper embrittlement of ultrahigh strength steels. *Int J Miner Metall Mater*. 2011;18:543–50. https://doi.org/10.1007/s12613-011-0475-0.

[21] Sitek W, Trzaska J, Gemechu WF. Modelling and analysis of the synergistic alloying elements effect on hardenability of steel. *Arch Foundry Eng*. 2022;22:102–8.

[22] McKinsey & Company. *The economic potential of generative AI: The next productivity frontier*. 2023.

[23] McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 1955. *AI Mag*. 2006;27(4):12. https://doi.org/10.1609/aimag.v27i4.1904

[24] *Encyclopedia Britannica*. Artificial intelligence. Available from: https://www.britannica.com/technology/artificial-intelligence [Accessed 24 Jan 2025].

[25] Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev*. 1959;3(3):210–29.

[26] Taye MM. Understanding of machine learning with deep learning: Architectures, workflow, applications and future directions. *Computers*. 2023;12(5):91. https://doi.org/10.3390/computers12050091

[27] IBM. Machine learning. Available from: https://www.ibm.com/think/topics/machine-learning [Accessed 24 Jan 2025].

[28] Schmidhuber J. Deep learning in neural networks: An overview. *Neural Netw*. 2015;61:85–117.

[29] Davidovits P. Electricity. In: *Physics in Biology and Medicine*. 5th ed. 2019. p. 193–211. https://doi.org/10.1016/b978-0-12-813716-1.00013-6

[30] Koo CL, Liew MJ, Mohamad MS, Salleh AH. A review for detecting gene-gene interactions using machine learning methods in genetic epidemiology. *Biomed Res Int*. 2013;2013:432375. https://doi.org/10.1155/2013/432375

[31] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys*. 1943;5:115–33.

[32] Teleanu RI, Niculescu A-G, Roza E, Vladâcenco O, Grumezescu AM, Teleanu DM. Neurotransmitters—Key factors in neurological and neurodegenerative disorders of the central nervous system. *Int J Mol Sci*. 2022;23:5954. https://doi.org/10.3390/ijms23115954

[33] Banerjee S, McCracken S, Hossain MF, Slaughter G. Electrochemical detection of neurotransmitters. *Biosensors*. 2020;10:101. https://doi.org/10.3390/bios10080101

[34] Abraham TH. (Physio)logical circuits: the intellectual origins of the McCulloch–Pitts neural networks. *J Hist Behav Sci*. 2002;38(1):3–25.

[35] Rosenblatt F. *The perceptron: A perceiving and recognizing automaton*. Ithaca (NY): Cornell Aeronautical Laboratory; 1957.

[36] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol Rev*. 1958;65(6):386–408.

[37] Bajohr H. The Gestalt of AI: Beyond the holism-atomism divide. *Interface Critique*. 2021;3:13–35. https://doi.org/10.11588/ic.2021.3.81304

[38] Rosenblatt F. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Buffalo (NY): Cornell Aeronautical Laboratory; 1961.

[39] Kolb H. How the retina works. *Am Sci*. 2003;91(1):28–35.

[40] Block HD. The perceptron: A model for brain functioning. *Rev Mod Phys*. 1962;34:123–35.

[41] Block HD, Knight BW Jr, Rosenblatt F. Analysis of a four-layer series-coupled perceptron. II. *Rev Mod Phys*. 1962;34:135–42.

[42] Block HD. Analysis of perceptrons. In: *Proceedings of the Western Joint IRE-AIEE-ACM Computer Conference*; 1961 May 9–11; San Francisco, CA. New York (NY): IRE; 1961.

[43] Van Der Malsburg C. Frank Rosenblatt: *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. In: Palm G, Aertsen A, editors. *Brain Theory*. Berlin, Heidelberg: Springer; 1986. (Proceedings of the First Trieste Meeting on Brain Theory). https://doi.org/10.1007/978-3-642-70911-1_20

[44] Hebb DO. *The Organization of Behavior*. New York (NY): Wiley; 1949. Hayek FA. *The Sensory Order*. Chicago (IL): University of Chicago Press; 1952.

[45] Olazaran M. A sociological study of the official history of the perceptrons controversy. *Soc Stud Sci*. 1996;26(3):611–59.

[46] Widrow B, Hoff ME. Adaptive switching circuits. In: *IRE WESCON Convention Record*; 1960; New York (NY): IRE. p. 96–104.

[47] Robbins H, Monro S. A stochastic approximation method. *Ann Math Stat*. 1951;22:400–7. https://doi.org/10.1214/aoms/1177729586

[48] Widrow B, Lehr MA. Artificial neural networks of the perceptron, MADALINE, and backpropagation family. In: Bothe HW, Samii M, Eckmiller R, editors. *Neurobionics: An Interdisciplinary Approach to Substitute Impaired Functions of the Human Nervous System*. Amsterdam: Elsevier; 1993. p. 133–205.

[49] Curtin CG, Allen TFH, editors. Non-linearity in natural, physical and social systems. In: *Complex Ecology: Foundational Perspectives on Dynamic Approaches to Ecology and Conservation*. Cambridge (UK): Cambridge University Press; 2018. p. 283–387.

[50] Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge (MA): MIT Press; 2016.

[51] Werbos PJ. *Beyond regression: New tools for prediction and analysis in the behavioral sciences* [PhD thesis]. Cambridge (MA): Harvard University; 1974.

[52] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, editors. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 1: Foundations. Cambridge (MA): Bradford Books/MIT Press; 1986. p. 318–62.

[53] Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: *Proceedings of the 27th International Conference on Machine Learning*; 2010 Jun 21–24; Haifa, Israel. 2010.

[54] Grossmann MA, Asimov M, Urban SF. Hardenability, its relation to quenching and some quantitative data. In: *Hardenability of Alloy Steels*. Cleveland (OH): ASM; 1939. p. 124–96.

[55] Grossmann MA. Hardenability calculated from chemical composition. *AIME Trans*. 1942;155:227–55.

[56] Crafts W, Lamont JL. *Hardenability and Steel Selection*. Toronto (ON): Pitman Publishing Corporation; 1949. p. 147–75.

[57] Kramer IR, Hafner RH, Toleman SL. Effect of sixteen alloying elements on hardenability of steel. *Trans AIME*. 1944;158:138–58.

[58] Comstock GF. The influence of titanium on the hardenability of steel. *AIME Trans*. 1945;1:148–50.

[59] Hodge JM, Orehoski MA. Relationship between the hardenability and percentage of carbon in some low alloy steels. *AIME Trans*. 1946;167:627–42.

[60] Just E. New formulas for calculating hardenability curves. *Met Prog*. 1969;75(5):87–8.

[61] Brown GT, James BA. The accurate measurement, calculation, and control of steel hardenability. *Metall Trans*. 1973;4:2245–56.

[62] Kunze CT, Russel JE. The generation of accurate Jominy data: its use to improve the Grossmann hardenability factors. In: Doane DV, Kirkaldy JS, editors. *Hardenability Concepts with Application to Steel*; 1977; Chicago, IL. p. 290–307.

[63] Doane DV. Application of hardenability concepts in heat treatment of steel. *J Heat Treating*. 1979;1:5–30. https://doi.org/10.1007/BF02833206

[64] Mangonon PL. Relative hardenabilities and interaction effects of Mo and V in 4330 alloy steel. *Metall Trans A*. 1982;13:319–20.

[65] Tartaglia JM, Wada T, Diesburg DE, Eldis GT. Influence of alloying elements on the annealability of carburizing steels. *J Met*. 1982;34(6):30–5.

[66] Kasuya T, Yurioka N. Carbon equivalent and multiplying factor for hardenability of steel. In: *Proceedings of the 72nd Annual AWS Meeting*; 1991 Apr 15–19; Detroit, MI.

[67] Yamada M, Yan L, Takaku R, Ohsaki S, Miki K, Kajikawa K, Azuma T. Effects of alloying elements on the hardenability, toughness and the resistance of stress corrosion cracking in 1 to 3 mass % Cr low alloy steel. *ISIJ Int*. 2014;54:240–7.

[68] Bhadeshia HKDH. Neural networks in materials science. *ISIJ Int*. 1999;39:966–79.

[69] Filetin T, Majetić D, Žmak I. Application of neural networks in predicting the steel properties. In: *Proceedings of the 10th International DAAAM Symposium*; 1999 Oct 21–23; Vienna, Austria.

[70] Dobrzanski LA, Sitek W. Application of a neural network in modelling of hardenability of constructional steels. *J Mater Process Technol*. 1998;78:59–66.

[71] Sitek W, Dobrzanski LA, Zacłona J. The modelling of high-speed steels' properties using neural networks. *J Mater Process Technol*. 2004;157–158:245–9.

[72] Sitek W, Trzaska J, Dobrzański LA. An artificial intelligence approach in designing new materials. *J Achiev Mater Manuf Eng*. 2006;17:277–80.

[73] Sitek W. Methodology of high-speed steels design using the artificial intelligence tools. *J Achiev Mater Manuf Eng*. 2010;39:115–60.

[74] Sitek W, Trzaska J. Practical aspects of the design and use of artificial neural networks in materials engineering. *Metals*. 2021;11:1832.

[75] Tomašić N, Sitek W, Iljkić D, Gemechu WF. Designing the chemical composition of steel with required hardenability using computational methods. *Metals*. 2024;14:1076. https://doi.org/10.3390/met14091076

[76] Vannucci M, Cola V. Automatic steel grades design for Jominy profile achievement through neural networks and genetic algorithms. *Neural Comput Appl*. 2021;33:16451–70. https://doi.org/10.1007/s00521-021-06242-w

[77] Geng X, et al. Prediction of hardenability curves for non-boron steels via a combined machine learning model. *Materials*. 2022;15:3127.

[78] ASTM International. *A255-10 Standard Test Methods for Determining Hardenability of Steel*. West Conshohocken (PA): ASTM International; 2010.

[79] Totten GE, Bates CE. *Handbook of Quenchants and Quenching Technology*. Materials Park (OH): ASM International; 1993. p. 35–68.

[80] Smoljan B, Iljkić D, Tomašić N. Mathematical modelling of hardness of quenched and tempered steel. *Arch Mater Sci Eng*. 2015;74:85–93.

[81] Massey FJ. The Kolmogorov–Smirnov test for goodness of fit. *J Am Stat Assoc*. 1951;46:68–78. https://doi.org/10.2307/2280095

[82] MathWorks. DeepLearning – trainlm. 2024. Available from: https://www.mathworks.com/help/deeplearning/ref/trainlm.html [Accessed 24 Jan 2025].

[83] Huang X, Cao H, Jia B. Optimization of Levenberg–Marquardt algorithm applied to nonlinear systems. *Processes*. 2023;11:1794. https://doi.org/10.3390/pr11061794

[84] Jierula A, Wang S, Oh T-M, Wang P. Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data. *Appl Sci*. 2021;11:2314. https://doi.org/10.3390/app11052314

[85] Zemlyak S, Gusarova O, Khromenkova G. Tools for correlation and regression analyses in estimating a functional relationship of digitalization factors. *Mathematics*. 2022;10:429. https://doi.org/10.3390/math10030429

[86] Emmert-Streib F, Dehmer M. Evaluation of regression models: model assessment, model selection and generalization error. *Mach Learn Knowl Extr*. 2019;1:521–51. https://doi.org/10.3390/make1010032

[87] Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci*. 2021;7:e623. https://doi.org/10.7717/peerj-cs.623

[88] MathWorks. DeepLearning – mapminmax. 2024. Available from: https://www.mathworks.com/help/deeplearning/ref/mapminmax.html [Accessed 24 Jan 2025].

[89] Xiong Z, Cui Y, Liu Z, Zhao Y, Hu M, Hu J. Evaluating explorative prediction power of machine learning algorithms for materials discovery using k-fold forward cross-validation. *Comput Mater Sci*. 2020;171:109203. https://doi.org/10.1016/j.commatsci.2019.109203

[90] Trzaska J, Sitek W. A hybrid method for calculating the chemical composition of steel with the required hardness after cooling from the austenitizing temperature. *Materials*. 2024;17:97. https://doi.org/10.3390/ma17010097

LIST OF FIGURES

LIST OF TABLES

BIOGRAPHY

Neven Tomašić, born on August 6, 1976, completed his secondary education at Sušačka Gimnazija before enrolling in the Faculty of Engineering at the University of Rijeka, where he graduated in 2001 with a univ. mag. ing. mech. degree.

Since 2002, he has been involved in scientific projects, including:

- *Computer-Aided Simulation and Development of Materials* (*Računalne simulacije i razvoj materijala*), principal investigator: T. Filetin,
- *Optimizing Parameters and Predicting the Results of Heat Treatment of Metals* (*Optimiranje parametara i predviđanje rezultata toplinske obrade metala*), principal investigator: B. Smoljan,
- *Optimization and Modeling of Thermal Processes of Materials* (*Optimiranje i modeliranje termalnih procesa materijala*), principal investigator: D. Iljkić.

These projects were funded by the Ministry of Science, Education, and Youth (formerly known as the Ministry of Science and Technology and the Ministry of Science, Education, and Sports) and the Croatian Science Foundation.

He has over 20 years of experience teaching in the STEM field, spanning elementary education, high school, and higher education. In addition, he has spent more than a decade working on innovative projects, including the successful transformation of industrial heritage and renewable energy projects.

He is the author of over 30 scientific papers in the field of materials science.

LIST OF PUBLICATIONS

List of the most relevant publications:

- Tomašić N, Sitek W, Iljkić D, Gemechu WF. Designing the chemical composition of steel with required hardenability using computational methods. *Metals*. 2024;14:1076.
- Megersa GK, Sitek W, Nowak AJ, Tomašić N. Investigation of the influence of fused deposition modeling 3D printing process parameters on tensile properties of polylactic acid parts using the Taguchi method. *Materials*. 2024;17:5951.
- Smoljan B, Iljkić D, Tomašić N. Mathematical modelling of hardness of quenched and tempered steel. *Arch Mater Sci Eng*. 2015;74:85–93.
- Smoljan B, Iljkić D, Tomašić N. Computer simulation of microstructure of quenched moulding die. *Arch Mater Sci Eng*. 2014;68:81–6.
- Smoljan B, Iljkić D, Tomašić N. Prediction of mechanical properties and microstructure composition of quenched and tempered steel. In: *Proceedings of the 28th ASM Heat Treating Society Conference "Heat Treat 2015"*; 2015. Materials Park (OH): ASM International. p. 266–73.
- Smoljan B, Iljkić D, Tomašić N. Computer simulation of mechanical properties of quenched and tempered steel specimen. *J Achiev Mater Manuf Eng*. 2010;40:155–9.
- Smoljan B, Smokvina Hanza S, Tomašić N, Iljkić D. Achievements of mathematical modelling of austenite decomposition during the heat treatment processes. In: *Proceedings of the 16th International Federation for Heat Treatment and Surface Engineering Congress*; 2007.